Australian Government
Department of Defence
Defence Science and
Technology Organisation

Structural Uncertainties in Numerical Induction Models

Lewis Warren

Command and Control Division
Defence Science and Technology Organisation

DSTO-TR-1895

ABSTRACT

This report delineates a number of ways in which the results of numerical induction models, which aggregate lower level measures into meta-measures for decision making, can be unnecessarily compromised. Examples of numerical induction models include complex models for performance evaluation, measures of effectiveness synthesis, and for strategic decision analysis. A framework is proposed for identifying different types of modelling uncertainty that may be present and several of these uncertainties are discussed in detail. Some popular decision analysis techniques are also analysed highlighting any features that may introduce unnecessary uncertainty into the results. The purpose of describing these potential pitfalls is to reduce the structural uncertainty forms that may be unwittingly added to the uncertainties that already exist in the input information leading to outputs that are more meaningful. More meaningful outputs should then naturally result in improved decisions when such models are applied to Defence problems.

RELEASE LIMITATION

Approved for public release

Published by

Command and Control Division
DSTO Defence Science and Technology Organisation
PO Box 1500
Edinburgh South Australia 5111 Australia

Telephone: (08) 8259 5555
Fax: (08) 8259 6567

© Commonwealth of Australia 2006
AR-013-699
July 2006

APPROVED FOR PUBLIC RELEASE

Structural Uncertainties in Numerical Induction Models

Executive Summary

A wide variety of conceptual models can be classed as numerical induction models that synthesise lower-level information into higher-level numerical information. In general, such induction models are used for strategic decision analysis, complex system reliability or vulnerability analysis, and cost/benefit/risk analysis. An important characteristic of these models is that their results cannot be definitively validated, and for this reason, it is important to be aware of the uncertainties that are inherent in the output numerical values. This report identifies a range of uncertainties that may be inherent in the cognitive structure of the model, as well as in the computational methods applied. Overall, this category of uncertainty can be termed "structural" uncertainty, in contrast to the intrinsic forms of uncertainty inherited by the information elements that a model processes.

Initially a schema of multiple levels of uncertainty is defined for the purpose of identifying the various mixtures of uncertainty in decision analysis techniques. Using this schema various forms of structural uncertainty in some common decision theoretics are identified, as well as the concomitant limitations they can impose on the modelling outputs. One common complication in complex problems, which prevents the use of linear information aggregation methods, is the presence of information interdependencies. For this reason, several types of interdependency that may exist between information elements relating to the different facets of a model are also described. As minimum requirements for the selection of an adequate information aggregation technique, some aggregation axioms are proposed. Aggregation operators are then discussed of two categories: those that are functions of individual information elements and those that are functions of sets of information. Operators of the second category are suggested to be more appropriate for capturing aggregation non-linearities that are present in sets of interdependent information.

With a clear awareness of the structural uncertainties that can unwittingly be invoked in a numerical induction model, an analyst may increase the value that is added to the input information and embed more meaning in the model outputs by choosing techniques which minimise these potential uncertainties. Improving the quality of such modelling outputs would then lead to improved solutions to the complex Defence problems being addressed by this type of model.

Author

Lewis Warren
Command and Control Division

Lewis Warren has a background in Operations Research, simulation, and mathematical modelling for business application development. His educational qualifications are B. Industrial Eng. (Hons) from University of Melbourne, Master of Engineering (Systems) from RMIT University, and a PhD in Complex System Modelling from Swinburne University of Technology. In recent years he has focused on soft OR and AI techniques, especially uncertainty representation and analytical techniques for strategic decision analysis and complex systems performance evaluation. While at DSTO he has worked in the field of Information Operations developing effects-based analytical processes and techniques to support full-spectrum Operations planning.

Contents

Acronyms

ABT:	Alternative-Based Thinking
AHP:	Analytic Hierarchy Process
ANN:	Artificial Neural Network
ANP:	Analytic Network Process
CEV:	Choquet Expected Value
EBO:	Effects-Based Operations
ELECTRE:	French decision theoretic of Roy
FCM:	Fuzzy Cognitive Mapping
MAUT:	Multi-Attribute Utility Theory
MCDM:	Multi-Criteria Decision Making
MOE:	Measure of Effectiveness
OWA:	Ordered Weighted Average
SD:	System Dynamics
TOPSIS:	Technique for Order Preference by Similarity to Ideal Solution
U:	Uncertainty
VFT:	Value-Focused Thinking
ZAPROS:	Decision theoretic of Larichev and Moshkovich

Glossary

Datum:	A reference point to measure from
Fuzzy measures:	Non-additive weights applied to sets of measures
Scale Type:	Categorical, Ordinal, Interval, or Ratio scales
Scale Unit:	Basic division of a scale for measurement
Exchange Ratio:	A parameter to convert units of one variable to those of another.

1. Introduction

Complex decision problems are often modelled by decomposing coherent aspects of the problem into successively smaller elements. This has been called the "divide and conquer" modelling approach. Analytical inputs are then assessments, quantitative or qualitative, of the value, performance, or some contribution of the elemental components to the overall problem or system. Subsequently, these elemental assessments need to be aggregated into higher level measures as are associated with the higher level concepts. This type of decision analysis may be used to compare different decision alternatives, or alternatively, the model may simply represent a complex system and the overall measure be used to assess some kind of behavioural performance. In Defence such models may be applied to a broad range of problems including capability evaluation, capability planning, military operations measures of effectiveness, system reliability, or performance of a system or system of systems.

A wide range of techniques have also been developed to facilitate the aggregation of the elemental information inputs, and several of the most common methods are discussed in this report. However, some techniques tacitly embed limiting assumptions and may also incorporate steps that are fanciful and difficult to mathematically justify. When important decisions are based on this type of model it would seem that the model and its techniques should be an adequate fit to the actual characteristics of the problem. Any misfit to the problem characteristics, or inexplicable operations, or unreasonable assumptions, would then increase the level of uncertainty surrounding the results.

The objective of this report is to raise the level of awareness of analysts to these different types of modelling uncertainty so that overall model uncertainty may be minimised as much as possible. To this end a conceptual framework and uncertainty typology is presented by which to identify the pitfalls which could compromise the value and usefulness of model results. This is particularly relevant to those Defence models which are composed of many abstract concepts and where there may be loose couplings as indeterminate inter-dependencies between elements. For convenience, the term "numerical induction" will be adopted to describe the general type of divide and conquer model addressed by this report.

Section 2 first provides a general overview of the inductive process. Section 3 describes various levels of modelling uncertainty and outlines the foundations for the typology of uncertainty sources that this report is based upon. Section 4 next examines the different structural forms of uncertainty resulting from the form of the conceptual model and different kinds of inter-relationships that can exist between its elements. Subsequently, Section 5 discusses various aspects of information aggregation in complex system models. Section 6 then illustrates how some structural types of uncertainty may be intrinsic to a decision analysis technique, and identifies some of these uncertainty forms in several common decision theoretics. Finally, Section 7 outlines some general areas where structural uncertainties may compromise the results of Defence analytical models.

2. Overview of the Inductive Process

Inductive processes map from particular instances to a general proposition. On the other hand, deductive processes map from general propositions to the particular, usually by feeding particular instances into rules, principles, or hypotheses, and deducing conclusions. In general, the term "induction" usually implies "logical" induction where the mapping proceeds through logical reasoning of one kind or another. An example of logical induction is where a general mathematical expression is derived for the sum of numerical series based on an observed pattern in a small part of the series. There are a number of logical induction variants and some of these will now be briefly summarised. These summaries have been extracted from [39] where a more detailed discussion of induction and its relevance to the scientific method can be found.

Naïve Induction
This is the oldest form of logical induction and it assumes that the process of gaining knowledge commences by collecting observations based on experience. Induction then proceeds from singular statements about particular instances to generalisations about all events of a certain kind at all places and times; as exemplified by the generation of scientific theories based on a compendium of observations. However, a number of criticisms have been levelled against naïve induction. Perhaps the major criticism is that with this somewhat ad hoc approach there is no guarantee that contradictory facts or exceptions will not appear sometime in the future. Thus, it can only result in a tentative level of truth, and it is said to be naïve because it relies heavily on the inductive method itself.

Sophisticated Induction (Logical Positivism)
This form of induction evolved from the latter half of the nineteenth century as a result of the previous criticisms. Its main assumption is that knowledge can be derived from experience as distinct from observation. As stated in [39]: "scientific theories and laws are not the simple summation of individual propositions. Rather, theories are an axiomatic network of statements from which singular propositions can be derived, and subsequently verified." This describes modifications to the fundamental structure of knowledge, and from this viewpoint knowledge is primarily a basis for making predictions. It is a more flexible reiterative procedure where induction results in a generalisation, a hypothesis is then proposed based on the generalisation, predictions are validated by experiments or observations, and modifications to the original generalisation (theory) are made if necessary by inductive inference. The main criticism of this approach is that the verification of a theory requires statements to be made such that they can be confirmed or otherwise. Unfortunately, abstract constructs and ideas cannot always be so stated, and in these cases, the reiterative procedure ceases and the method falters.

Popper's Falsification
Karl Popper around 1959 introduced a variant to the above Logical Positivism, known as the inductive-hypothetico-deductive method. Rather than seeking constant *affirmation* of a theory, the objective is to constantly test or expose theories to *refutation* in order to cull the less robust ones. From this viewpoint no theory can be considered to be absolutely true, and the type of observations that are required are determined by the theory itself rather than by a hypothesis.

The objective of falsification is to reduce ambiguity in scientific knowledge by elimination rather than inclusion, and the growth of knowledge is an evolutionary process where new theories constantly absorb old theories. From this viewpoint, observations are considered to be fallible and cannot be relied upon to confirm theories. However, one problem with this approach is that no theory can be absolutely falsified because of the effect of auxiliary assumptions, as well as the effect of variations in initial conditions (a problem also identified in chaos theory).

Post Positivism (Current Inductive-Hypothetico-Deductive Method)
The current state of the scientific method can be considered as a loose mix of tenets and principles of scientific inquiry. It is a pluralism of ideas under which there is no coherent methodology. Nevertheless, theory development continues to be the foundation of the scientific method without which no research objectives could be achieved.

The main purpose of introducing the brief descriptions above of the scientific application of logical induction, is to indicate that the results of applying the inductive process are generally associated with various kinds of uncertainty, and that the intent should be to minimise the uncertainties within the modelling process as much as possible because the results cannot be absolutely validated. Some examples of logical induction that can be found in the field of data mining are algorithms used for decision tree or rule extraction (induction) from large amounts of data. Another example is the application of a backward induction process in a sequential game (or decision tree) to find one (but not all) equilibrium path for determining optimal decisions at nodes within the acyclic decision graph.

Numerical Induction
With respect to the numerical integration of behavioural measures, the psychologist Anderson [5] in 1981 noted :
"*Integration theory has operated primarily in the inductive mode whereby generalisations are sought as emergents from experimental analysis*" and also,
"*Inductive theory views science not as formalised knowledge, but as living enquiry.*"

In a similar spirit this report will apply the term "numerical induction" to mathematical models which synthesise sets of numerical measures into more global aggregate values for decision making. This type of measure aggregation will be considered as an inductive process meaning from the many to the singular. Frequently, such inductive processes proceed through multiple levels guided by the conceptual model, rather than by logical reasoning. However, the distinction between deductive and inductive processes is not always so clear, especially in models where measures are propagated across a model.

Some models may possess degrees of both deduction and induction. For example, the so-called "backward induction" process sometimes applied to sequential games [1] could be viewed as a deductive process because it is driven by a set of propositions, or rules, which guide the selection of a path from many alternatives. Thus, the procedure for decomposing the global measure (maximum gain) into its component measures is not simply numerical induction reversed. And as there are always doubts about the absolute validity of the results of logical induction, there may also be doubts associated with the results of any numerical induction model. In this case, the doubts result from the imperfect nature of the cognitive models which are, after all, only simplifications of the world. The relative degree to which a model simplification reflects the real problem can be viewed as a function of the following factors:

- The finesse with which the key features and characteristics of the problem are identified.
- The nature of the conceptual variables defined to characterise the problem.
- The completeness of the set of conceptual variables.
- The assumptions that the structure of the model is based upon.
- The structure of the conceptual model itself.
- The information synthesis mechanisms and their assumptions.
- The finesse of uncertainty representation.
- The adequacy of uncertainty management within the induction process.

Inadequacy or a deficiency in any one of these factors may severely limit the meaning of the induced numerical values to the extent that they may not be able to provide any useful information about the real-world problem. Anderson [5] has also discussed in some detail various problematic issues confronting the integration of information in numerical induction models, as have some other authors also [39,60]. For this reason, this report aims to highlight how deficiencies in these areas can unwittingly be introduced into many common numerical induction applications. Even the most obvious deficiency of applying an inappropriate model to a problem, where the model features do not match the problem characteristics, is not infrequently encountered. Moreover, there are a variety of other more subtle pitfalls that can decrease the usefulness of numerical outputs and these potentially limiting factors may not appear to be very important because they are so basic and fundamental. Increasing the mathematical and algorithmic complexity of a model is also frequently assumed to be the best way to increase the usefulness of a model.

For these reasons, this report discusses a range of fundamental considerations that can impact on the selection of conceptual modelling methods, such as the basic cognitive form and information synthesis methods.

3. Levels of Uncertainty in Induction Models

The adoption of a conceptual model for computational analysis of a problem imputes several levels of potential uncertainty (U). These uncertainty levels fall into two main categories: uncertainty forms within information elements, and uncertainty forms in the conceptual model and its computational techniques, which are here termed the *structural uncertainties* of the model. The various forms of uncertainty that may be present within information elements are outside the scope of this report, but are examined in an accompanying report [93].

This report will propose and adopt the following framework or typology of U levels for the purpose of identifying the various types of U that may be present in a model. From these definitions for levels of U, higher numerical values actually represent lower levels or more granular forms of uncertainty. As stated above, the U levels can be grouped into two main classes: the first class is associated with the inherent model structure and its mechanisms (Levels 1-4), and the second class concerns the U introduced by the sets of data/information that is processed by the model (Levels 5-6).

3.1 Levels of Modelling Uncertainty

Level 1: Uncertainty in Objective or Problem Definition --
Uncertainty of the purpose of the analysis may be related to mistaken perceptions, confusion, lack of information, or complexity. Personal factors such as experience, skill and bias can influence the cognisance of a problem and some analytical methods such as personal construct theory, cognitive mapping or the Soft Systems Methodology, may assist in consolidating the problem definition and variable identification. This level of U may also be invoked when a concept used to describe a problem is given a different interpretation by different individuals.

Level 2: Uncertainty in Model Conceptualisation --
At this level a conceptual model is adopted as the computational framework. A range of questions must be answered to define the broad characteristics of the model and fit it to reality in an adequate manner. Such considerations are:
What paradigm : Single or multiple models? Random variables or not?
What structure: Where are the boundaries? How detailed and what model granularity? Hierarchical or complex system inductive model?

Uncertainty can arise at this level if the choices made to answer these questions are inept and do not fit the problem characteristics with an adequate degree of verisimilitude. An example of U introduced at this level is when a probabilistic inferencing model is adopted where the characteristics of the problem would suggest that an inductive type information integration model would be more appropriate for decision making. Although it may not always be clear as to what represents an "adequate" fit, and because there may also be multiple fits that are adequate, there can still be cases of greatly oversimplified models based on unreasonable assumptions as will be discussed in Section 3.

Level 3: Uncertainty in Computational Macro-Structure --
This level refers to the U introduced by major computational components of a model such as the inferencing or clustering techniques, the type of nodal squashing function in a neural net, or the information aggregation procedures in numerical induction models. Section 5 will discuss non-linear information aggregation requirements for certain types of systemic complexities, while Section 6 will identify U forms that are introduced through the methods used in some popular decision analysis theoretics.

Level 4: Uncertainty in Computational Micro-Structure and Parameters–
This level refers to the U introduced by analytical method components such as arbitrary parameters as in squashing function gain and bias in the fuzzy cognitive mapping technique, aggregation optimism/pessimism parameters in some aggregation operators, or the prior conditional probabilities in Bayes Nets. Some U implications of arbitrary parameters are also discussed in Section 6 in relation to normative decision theoretic models.

3.2 Levels of Information Uncertainty

Two additional levels of U pertain to the individual information elements themselves and the set of information elements available.

Level 5: Uncertainty in Sample Evidence --
Quantity - the amount of information affects measure estimation.
Quality - conflicting information in a sample also affects estimation.

Level 6: Intrinsic Uncertainty within Information Elements--
This is U inherited from variable definition and/or measurement.
A qualitative concept associated with a variable may be inherently vague, and measurements pertaining to any type of variable may be approximate, subjective, or indirect.

The above six levels of U are the broadest categories of U and have been defined in this report to facilitate the identification of the different sources of U in a model. The levels of U which are the primary focus of this report are Levels 1 to 4 which cover U forms within the structure of a model and its computational techniques.

Level 5 U (the collection of data elements) and Level 6 U (due to conceptualisation and measurement aspects of variables in a model) are forms of U induced by elemental information aspects. These two levels of information U are addressed in some detail in an accompanying report [93] where a new approach is presented for representing hybrid combinations of U in sets of information elements, as well as methods for measuring the aggregate U of the hybrid U combinations. Applications of the proposed methods are also demonstrated in that report. Vague variable definition and/or inexact measurement of a variable's property may lead to what will be called "soft" information in this report. This term does not include probability estimates derived from a statistically sufficient sample of data, but does include subjective probability estimates derived from a statistically insufficient sample of data.

4. Structural Uncertainty in Models

4.1 The Problem Definition

Broad conceptual models as abstract approximations of reality provide the frameworks for integration operations in numerical induction computations. However, the fitness of a model to the problem characteristics can greatly influence the quality of the output [97]. For this reason, the mental model adopted should be such that it enables the complexities of the real-world problem to be captured and adequately addressed in computations. Generally speaking, models may be formulated by human assignment or by automatic data analysis. The success of both approaches to model formulation depends on the choice of suitable knowledge representation schemas. The choice of concepts for model variables is also of fundamental importance. Categorisation techniques [65] may be used to address the problem of determining the "best" degree of detail (or abstraction) to yield the most useful model outputs. On occasion, small world, local problem models are also extended to approximate large world problems when the dimensions of those problems are unknown [53]. In recent times, attention has also been directed [20] towards the automated formulation of models such as decision tree induction, which is often based on the notion of Occam's Razor whereby the simplest explanatory representation is the best. However, many difficulties prevail and the validity of Occam's assumption has also been questioned [95]. This section will focus only on human assigned models, where the key problem is to define a numerical induction model and its computational mechanisms to capture to a sufficient degree the real-world complexities in the problem domain.

4.2 Conceptual Models for Information Synthesis

Several types of models will now be described that may be used to transform input information into output numerical values of interest. A loose interpretation of numerical induction is adopted for some of these models which do not necessarily transform inputs into more generalised information, but rather update state values within the model.

4.2.1 Networked Nodal Models

Network type models may propagate state changes in variables through a web of associative influences between variables. Alternatively, the nodes may not have attributes with state values and only be routing points for flows of information as in communication system models.

4.2.1.1 Acyclic Models

This type of network model does not allow feedback loops so that influence links are uni-directional between parent and children nodes. Thus, the possible associations are constrained.

<u>Examples</u>: A Bayes Net where the nodes represent random variables which are related to parent nodes by sets of conditional probabilities. The nodal states are thus probabilistically inferred.

An Artificial Neural Net (ANN) where an input state vector is transformed into an output state vector through layers of intermediate nodes, by passing weighted sums of input activations through non-linear squashing functions at each node. The ANN learning mode of operation recursively iterates through both directions to optimise weights for given input/output data sets.

4.2.1.2 Cyclic Models

In this type of network model feedback loops (reinforcing or inhibiting) are allowed enabling complex system behaviour to be emulated. A temporal dimension is necessary to explore the dynamics of the complex behaviour, which may tend to chaotic, fixed cycle or fixed point patterns. Frequently, such models are only used to estimate states of key variables in the near future. These models are not very suitable for the form of soft information synthesis that is the subject of this report. When they are applied to such information synthesis problems, the aggregate value is sometimes considered to be the steady-state value of a key node after perturbations that were initiated by an input set of information have ceased.

<u>Examples</u>:

- System Dynamics [83] models where key variables are quantitative "level" variables that are influenced by exchange rates, which may in turn be influenced by binary state qualitative variables. The use of qualitative variables within these models is rather restricted since the primary propagation technique is through differential equations and the division operation between *different* qualitative variables whose measures are on interval scales is not admissible.
  (This will be discussed further in Section 5.3.3.)

- Fuzzy Cognitive Maps [49] are cyclic models where variables are qualitative so these models are suitable for modelling the dynamic behaviour of abstract strategic situations.

- Saaty's [76] proposed Analytic Network Process (ANP) for decision making with interdependencies is another example of a cyclic network model.

4.2.2 Multi-Tier to Global Variable Models

This type of model is closer to a numerical induction model because its function is to condense low-level information into more generalised global forms of information. No feedback is allowed and parent-child influences are uni-directional.

4.2.2.1 Hierarchical Models

This is the most common type of model used for decision analysis. Strictly independent lower level information is hierarchically aggregated in a linear manner into higher-level decision values commonly using weighted additive operations.

<u>Examples</u>:

Many multi-attribute decision techniques including the Analytic Hierarchy Process [74] and probabilistic decision trees.

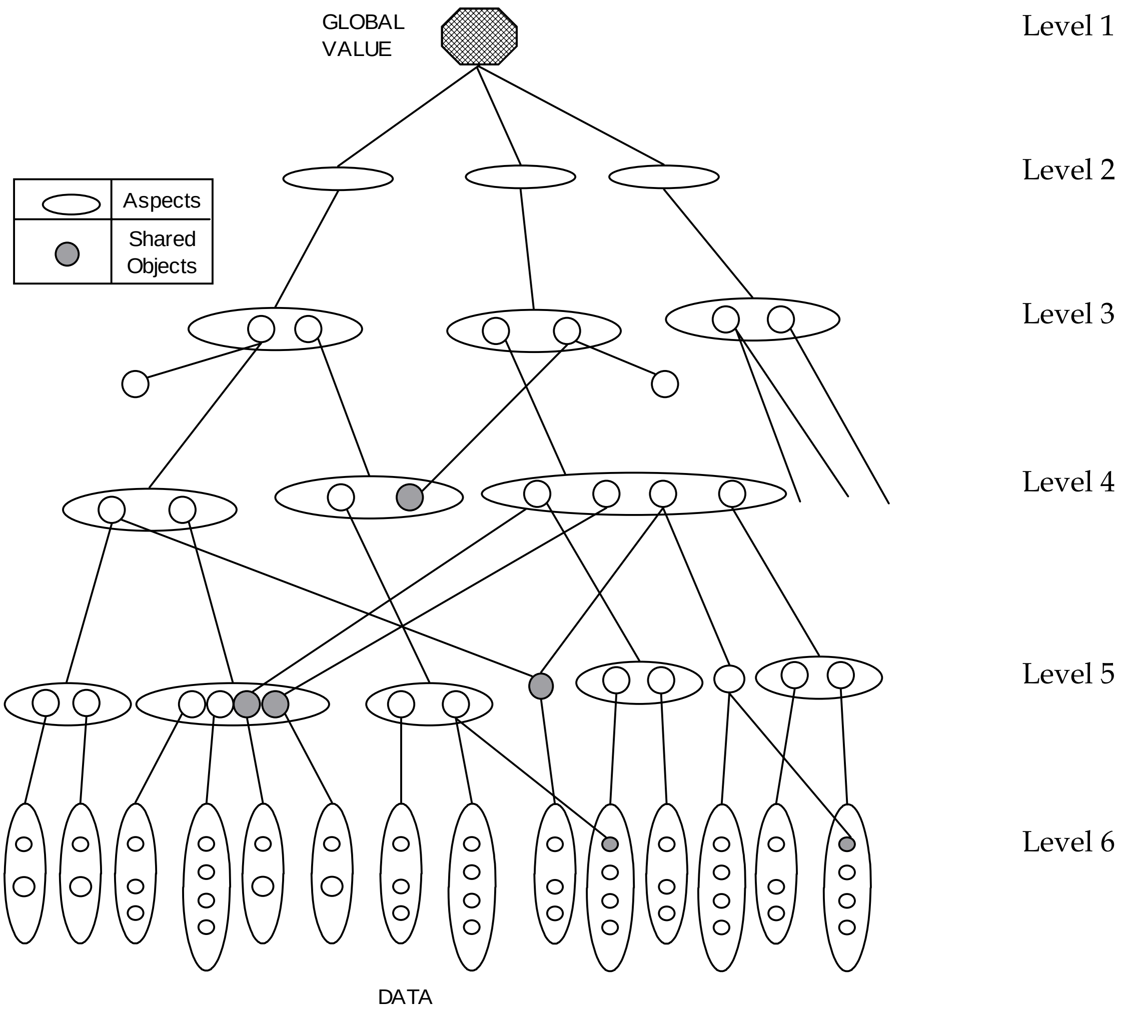

***Figure 6**: A Multi-Tier Model of Diverse Decision Facets*

4.2.2.2 Non-Hierarchical Models

This type of multi-tier model does not require strict independence of elements within or between levels, and different child nodes can share the same parent node. Such non-hierarchical information structures are particularly suitable for modelling abstract concept granulation and composition, as required in the analysis of complex domains or decisions, by the successive combination of diverse lower level information between which indeterminate interdependencies may exist. In Figure 6, the loosely coupled elements within the different facets or subsystems are grouped within the ellipses.

Examples: A special type of Bayes Net where multiple nodes condense into a small number of higher level nodes. As previously noted, care must be taken in such probabilistic models that the higher level nodes are true random variables and are not concepts that are deterministically induced from lower level node states.

Non-hierarchical models may also consist solely of deterministic variables (without probabilistic inferencing) and in this case the intrinsic nature of the variables determines what information aggregation procedures are appropriate. If the variables represent stand-alone entities that are not merely abstractions of lower level information, then weighted aggregation may be feasible whereby weights represent the strength of the influence of a parent on a child. However, additive weighted aggregation would assume that the parents are all independent, which is not true in fact because some may be influenced by common elements due to cross influences. Any U introduced by assuming independence would be Level 3 U concerning the computational macro-structure, and whether or not this would produce very misleading results would depend on the characteristics of the individual problem.

If the variables are not stand-alone concepts, but are abstractions of lower level information as are many decision analysis models, special non-additive information synthesis procedures may be required to capture the forms of dependency that may exist between the abstract concepts (as will be discussed further in Section 4.3). It should also be noted that the multiplicative aggregation technique that has been proposed by Keeney and Raiffa [45], although not additive, does not address the type of non-linearities due to interdependencies that will subsequently be described. Simply speaking, multiplicative models can be described as "not-additive" rather than "non-additive". In other words, their primary purpose is not to capture non-compensatory aspects, but rather to capture implicit multiplicative relationships as found in probabilistic laws for example.

4.2.3 Geometric Models

It will now be suggested that for understanding multi-tier models with interdependencies, the metaphor of composite geometrical structures may facilitate the visualisation of non-linear interactions in measure aggregation. One example of such geometric metaphors are complex globular bubbles as in Figure 7. Whereas multi-tier models help to identify where cross inheritance relationships exist, geometric models can help to get a feel for the non-linear quantitative relationships between measures. Although it may not be possible to visualise the whole problem in such a way, such a conceptualisation may help to understand the sub-component non-linearities. For example, the composite bubbles in Figure 7 could represent global variables whose surface area can be observed to be a non-linear function of the different component bubbles whose surface area represents their individual measures.

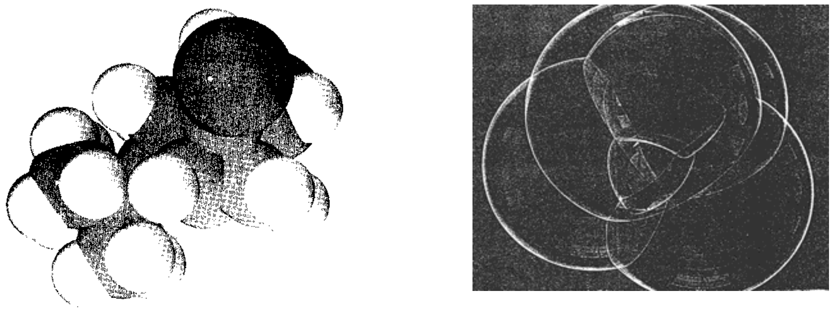

Figure 7: *Globular Geometric Models*

4.3 Interdependency Between Information Elements

Four potential types of information element interdependency will now be described which would indicate that non-linear information aggregation methods are required in the numerical induction process.

4.3.1 Causal Interdependency

Causal influence is the most familiar form of interdependency where the value(s), or state(s) of one or more elements, determines to some degree the value of the dependent element. These may be due to physical or abstract relationships between conceptual entities. Network models are frequently used to process information through such a network of relationships in a step-wise behavioural analysis, and learning algorithms applied with input/output training sets to determine causal link strengths. Alternatively, link strengths can be input initially and used to determine state values of element nodes. Network models with feedback can also be used to understand the dynamic behaviour of a system. One dynamic causal influence modelling technique is Fuzzy Cognitive Mapping [49] which is suitable for models with qualitative concepts. Another dynamic causal influence modelling technique is System Dynamics which is suitable for more quantitative concepts in a model.

4.3.2 Systemic Interdependency

This type of interdependency may exist between elements, or groups of elements, for a variety of reasons. But whatever the reason, the combination of information elements or subsets of the power set, is non-additive i.e. the integrated sum does not equal the sum (or average) of the parts. Systemic information sets may relate to tangible or abstract systems. An example of a tangible system is a bicycle with the subsystems: frame, brakes, wheels, gears and accessories. An example of an abstract system is the representation of a cost, benefit, risk problem; all being interdependent. For systemic information sets, non-compensatory aggregation (to some degree) is also desirable because low values of one system aspect cannot be compensated for by higher values of other system aspects, as with additive averaging or aggregation. The reason for this is that *all* aspects (or components) need to be considered, since the system consists of the totality or integration of related components (i.e. all components are indispensable). In other words, low values represent windows of weakness that can jeopardise the behaviour of the complete system. For example, weak brakes endanger the performance of a bike when it goes downhill because they cannot be compensated for by a very good seat. For this reason, when there are systemic interdependencies information aggregation should also reflect the *amount of disparity* within the set of measures, and non-compensatory methods should be applied to synthesise a global value.

4.3.3 Associative Interdependency

Associative interdependency can also exist when there is synergy or redundancy between sets of model elements and their measures. In these cases non-additive aggregation is also required. One type of associative dependency is also induced when a set of information elements are all conditionally influenced by one other factor. This can be called conditional independence where the elements are not related themselves, but exhibit some correlated behaviour due to all being conditioned by another variable. Similarly, in induction models when a meta-variable measure is to be induced from the set of information elements, non-compensatory integration is also desirable because the meta-value should again be indicative of the inconsistency or degree of disparity of the information set. Accordingly, the meta-variable measure should be decreased with increasing sub-element inconsistency to conform to the Principle of Maximum Uncertainty (and conversely the Principle of Maximum Information Entropy). An example of associative dependency is where a global performance measure is to be induced for a *group* of people from a set of measures pertaining to various individual behavioural characteristics. Such groups of people could be a division of a military combat force, or a group of workers in an organisation, and individual behavioural characteristics may relate to morale, abnormal actions, accidents, and goal achievement. In such examples, the meta-variable may represent the *group performance measure,* which is quite different from a *task performance measure* that would most likely be based on a set of efficiency and effectiveness measures.

4.3.4 Cognitive Interdependency

The fourth type of interdependency will be termed here "cognitive interdependency", induced when the synthesis of a set of numerical measures aims to emulate the human cognitive process. Various behavioural researchers, for example [5], have determined that the human assimilation of a set of numbers does not usually occur as with probabilistic type averaging. Rather, human evaluation is by searching for a dominant value or the simplest pattern in the set. In this way, cognitive interpretation is also influenced by disparities and similarity patterns within the numerical set. Zeleny [102,103] has introduced the concept of "cognitive equilibrium" for this type of non-compensatory and non-additive assimilation of numerical information. One situation where this type of information integration is appropriate is the synthesis of diverse expert opinions where an average (simple or weighted) may be very different to many opinions because of the effect of one large outlier. Besides emulating human cognition, another reason for non-compensatory aggregation of expert opinion is that, again, it is desirable to embed information on the disparities present, because the degree to which opposing views are present (discord and polarisation) is meaningful and should somehow be reflected *in* the global value. This type of synthesis cannot be achieved by using a variation statistic, nor through any type of generalised compensatory weighted averaging.

4.4 Section Summary

This section has discussed different types of conceptual models and the various types of interrelationships and interdependencies that may be present between the components of a conceptual model used for numerical induction. It has also been proposed that for the non-causal types of interdependencies, geometrical models may help to conceptualise the non-additivity required in the numerical induction process. The following section will present a survey of information aggregation procedures including non-additive methods. However, modelling the dynamic type of causal dependency is beyond the scope of this report.

5. Information Aggregation Techniques

5.1 Overview of Unconstrained Information Synthesis Techniques

A summary of different classes of unconstrained information synthesis is presented in Figure 8. In this context, the term "unconstrained" is used to differentiate these aggregation methods from mathematical programming and other quantitative methods which integrate information to determine (optimal) values of decision variables subject to constraints. While "synthesis" is here used here as an umbrella term for *any* method of combining discrete information, "aggregation" will specifically refer to the derivation of global *measures* for variables which may be of qualitative or quantitative definition. Global aggregate measures may then represent the property of decision value or utility, or be a measure of a decision maker's preference (sometimes called "priority"). Information synthesis may also be performed to rank order (partial or complete) a set of alternatives or choices without developing any global measures. This is referred to as the Global Order Vector in Figure 8. A further type of information synthesis shown in Figure 8 is the propagation of evidence through a model as it appears, to update belief values across its state variables. Although this type of information synthesis is beyond the scope of this report on numerical induction, a wide range of graph-based techniques has been proposed for it [4,64], plus non-graphical methods such as the use of the Sugeno fuzzy integral [86].

An important consideration for any type of information synthesis is whether there is any synergy or redundancy in the set of information, both being forms of information interdependency. Figure 8 includes examples of methods suitable for information synthesis when dependencies are present, as well as methods that are only suitable for independent information. Several of these methods will also be described further in Section 6. Furthermore, when interdependencies are present, it is important to identify what types of interdependencies are present so that an appropriate technique may be selected. Unfortunately, it is not uncommon in the literature for overzealous authors to imply that one particular method for modelling dependency suits all forms of dependency. This is simply not realistic or possible.

5.1.1 Additive Aggregation with Independent Information

For variables that are measurable, additivity is defined as: $\mu(A \cup B) = \mu(A) + \mu(B)$,
where A and B are disjoint sets of real numbers and μ is a measure.

Most traditional decision analytic techniques assume information independence and additivity, with an expected utility being computed using some form of a generalized mean; for example: arithmetic mean, geometric mean, harmonic mean, or one of various families of compensatory operators such as the Ordered Weighted Average [99]. All these aggregate measures fall in the range [Min, Max]. Some compensatory operators [105] have been used to model bias in utility aggregation which can be introduced by a decision maker's optimistic or pessimistic character; in other words, for modelling non-linearities in human cognition. This type of dependency is quite distinct to information interdependency. One extension to additive mean operators that has many variants is to fuzzify decision variables expressing variable values and weights as fuzzy sets. Generally speaking, fuzzy decision theoretics (as well as some others such as ZAPROS [51] and outranking methods [71]) attempt to capture the U inherent in the subjective ratings of the decision maker.

5.1.2 Non-additive Aggregation with Information Interdependencies

As described in Section 4.3 several types of interdependency in a numerical induction model may require special information synthesis procedures. Some examples of such techniques that have been proposed in the literature are also shown in Figure 8.

For example, associative dependency is present when different combinations of elements have special interactions. There may be synergy where the importance of the union is greater that the sum of the subsets, or there may be redundancy where the aggregate weight is less than the sum of the subsets due to an overlap of information content. There are also two types of associative non-additivity: a constant monotonic type of non-additivity which can be modelled by the g(λ) measure of Sugeno [86], or a non-monotonic form whereby subsets of the power set have composite weights that may not be steadily increasing in addition to being non-additive. To implement the non-linear or non-monotonic form of non-additivity all weight sets must be known either as input, or derived from analysis of a sufficient data set, which also requires some assumptions to be made. In general, the functionals called "fuzzy integrals" are useful for the associative type of non-additive information aggregation, as well as for the systemic and cognitive forms of interdependency.

Another type of utility interdependency is also shown in Figure 8, which is interdependency between the high-level objectives of the problem analysis. This can be viewed as systemic interdependency. Carlsson and Fuller [22] have proposed a method to address this form of dependency by establishing fuzzy tradeoff functions between the objectives and determining an optimal global value that "satisfices" objectives based on a T-Norm conjunctive operation on the tradeoff functions. However, this approach requires the definition of a functional relationship between objectives and this would be difficult to realise in many situations.

Besides the global utility forms of information aggregation, the other major form of information aggregation shown in Figure 8 relates to synthesis of the decision maker's *preference* information. This type of information synthesis is used in many decision theoretics based-on decision makers' subjective preference ratings between alternatives, and across all criteria or facets of a problem to combine them into a global set of preference measures for alternatives. There are a variety of difficulties confronting methods used to synthesise global preference measures and these will be discussed in Section 6. Theoretical examinations of generalised non-compensatory and non-additive preference structures can be found in [18,19,98].

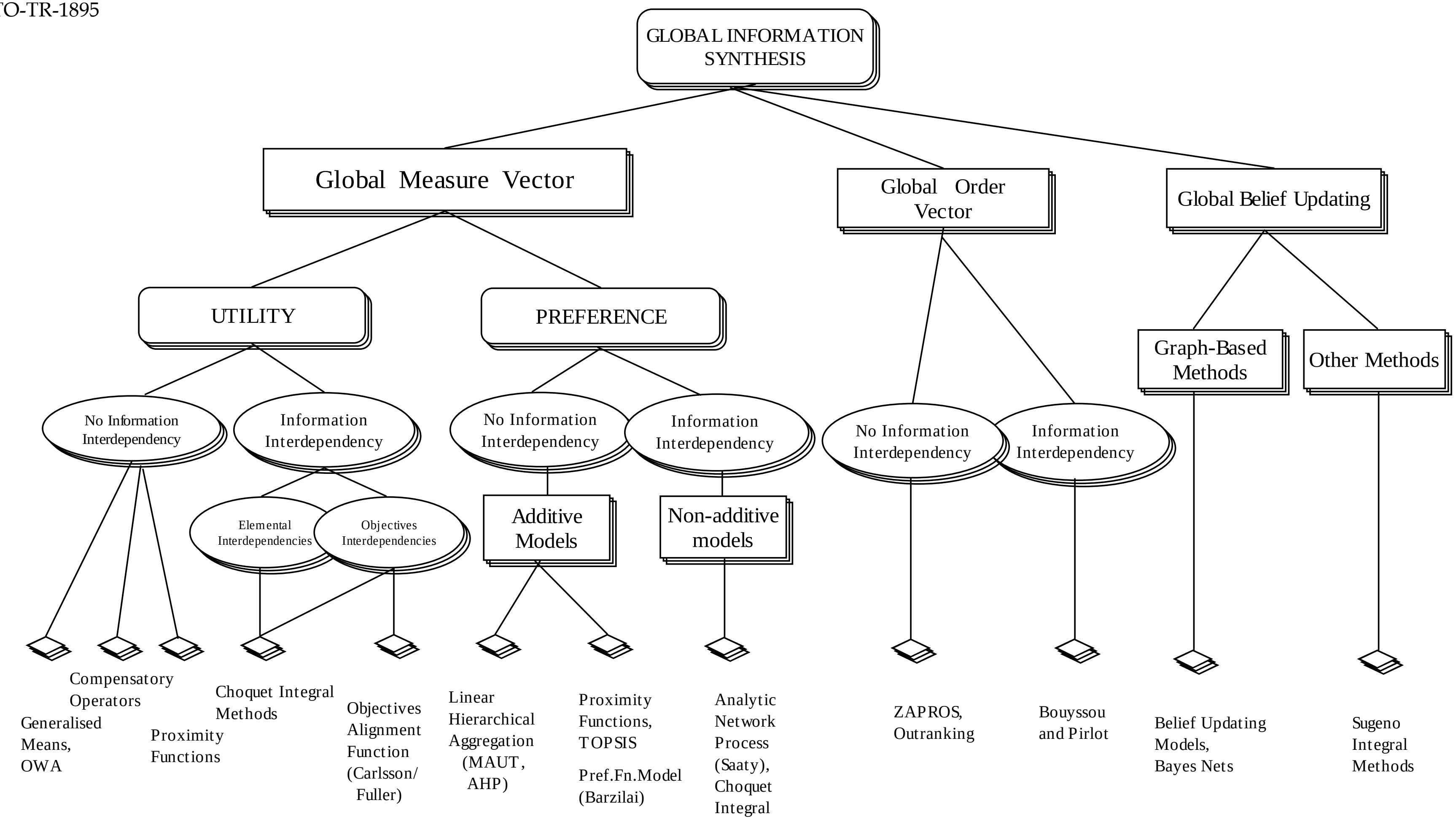

Figure 8: *Classes of Unconstrained Information Synthesis*

5.2 Overview of Information Aggregation Operators

This section will discuss some issues concerning aggregation operators as used to synthesise elemental information into global measures.

5.2.1 Some Information Aggregation Axioms

Since aggregation in this report has been defined to mean the development of a global *measure* from elemental information, it may involve other mechanisms besides simple summation. And although additive weighted summation is probably the oldest and most common form of information aggregation, it is not without some dangers that can limit the value and usefulness of its results. In order to avoid these dangers, some practical aggregation axioms will be defined for the meaningful *aggregation* of information measures. In the following paragraph "scale type" will refer to the different kinds of scale as discussed in Measure Theory [84]: categorical, ordinal, interval, and ratio scales (also discussed in more detail in Section 6.3). On the other hand, "scale unit" will refer to the basic divisions on a scale to which measures apply.

In general:

- Measures should have an equivalent scale type i.e. ratio or interval scales.
- Measures must have compatible scale units: e.g. uniform linear or logarithmic.
- Operations for the aggregation of measures are limited to the admissible operations for the scale type.
- Any synergy, as positive reinforcement or negative redundancy, existing between measures in a model indicates that non-additive information aggregation procedures are required.

It should be noted that there are exceptions to some of these axioms within some techniques. For example, it is admissible to multiply an interval scale measure by a ratio scale measure (i.e. different scale types) because that is not an aggregation operation.

5.2.2 Classes of Aggregation Operators

A very large variety of aggregation operators have been proposed for the combination of information in a model. Summaries of these can be found in [16,29,31,45,47,105]. To aid aggregation operator selection, Bloch [16] provides a classification for operators commonly encountered in the field of information fusion based on the following behavioural characteristics:

- Context independent constant behaviour
- Context independent variable behaviour
- Context dependent

These classes are further subdivided on the basis of partial ordering relationships, plus some mathematical properties that may be related to the real world problem. For example, *idempotence* can mean that repeated measuring of already known information will not change the previously derived aggregate, and *nilpotence* can mean that the accumulation of "n" pieces of information can lead to the null element as with cascaded opinion passing which depletes the value of the message. Other authors [29,98] note that aggregation operators may be classified in another way: conjunctive, averaging (including symmetric sums) and disjunctive. However, none of these rather general properties and categories are very helpful for identifying a suitable operator to address the previously described interdependencies that may be present in a decision problem. For this purpose, two different classes of operator will be identified: those which are functions of individual measures, and those which are functions of subsets of measures.

5.2.2.1 Operators as Functions of Individual Elements

Most operators that have been proposed are functions of individual measures. Although some are non-linear and non-compensatory to a degree, the specification of the non-linearity is rather arbitrary. For example, the so-called Zimmermann/Zysno general Compensatory AND operator [105] (which is actually partially compensatory), is a parametric function combining probabilistic OR (disjunction) that provides the maximum degree of compensation, with probabilistic AND (conjunction) that provides the minimum degree of compensation. These two expressions are linked by multiplication with a parameter (γ) that determines the degree of bias towards the OR direction of maximum compensation.

$$h(s_{1,} s_2 .. s_n) = \left(\prod_{i=1}^{n} s_i \right)^{1-\gamma} \left(1 - \prod_{i=1}^{n} (1 - s_i) \right)^{\gamma} \quad \text{where } 0 \le (s, \gamma) \le 1 \qquad (1)$$

While this operator could possibly be justified for modelling the personality bias of a known decision maker, it is of little use for modelling the complexities in systemic information, even though it is partially compensatory. The same criticism applies to most of the other non-additive or partially additive [98] operators, which are functions of individual elements since they cannot be related to the non-linearities that are functions of *groups* of elements as previously described. Nevertheless, the features of certain problems may sometimes point clearly to one of these as an appropriate operator.

5.2.2.2 Operators as Functions of Sets of Elements

Among the many aggregation operators those which are functions of variables that are groups of elements are rare. However, there is one operator among the so-called fuzzy integrals that aggregates subsets of information across the power set of elements: the Choquet capacity or Choquet integral [25]. In recent years there has been much interest in this operator for multi-criteria decision making [25,36,37,59]. This functional has the potential to capture most of the non-linearities described previously, although there can be some difficulty with its implementation, because the power set of weights of the subsets of elements is usually required which can be quite large. If the set of information is small, it may be feasible to subjectively evaluate the small power set of weights for some types of problem (as in [37] for a small set of student exam marks for a few subjects). For this reason, a large proportion of the literature on the application of the Choquet integral to decision making has focused on the problem of determining, or even optimising, the power set of weights from preliminary data analysis. This approach of searching for a unique power set of non-additive weights for a set of information not only requires a large data set for training, but also additional assumptions. However, in most abstract information synthesis problems, known and validated global aggregate values would seldom be available for training since the domain is generally hypothetical with aggregates that are impossible to validate. Thus, a different approach is required if the Choquet integral is to be applied for the synthesis of systemic information. This author has previously presented such an alternative approach to applying the Choquet integral in [91].

The Choquet Integral has been defined [36] in the following manner for a discrete and finite measure space. Let (X,χ,μ) be a fuzzy measure space where μ is the fuzzy measure of the power set χ, in the finite space X of all set elements. Sugeno [86] introduced the term "fuzzy measure" to describe non-additive weights for subsets of data as opposed to weights for individual data elements.

Consider a function $f: X \rightarrow [0,1]$ where $1 \geq f(x_1) \geq f(x_2) \geq f(x_3) \ldots \geq f(x_n) \geq 0$. The Choquet integral (C) of the function f with respect to fuzzy measure μ is defined by,

$$C\ (f(x_1),\ldots.f(x_n)) = \sum_{i=1}^{n} \left(f(x_i) - f(x_{i+1})\right)\mu(A_i) \qquad (2)$$

where $\mu(A_n) = 1$, and $f(x_{n+1}) = 0$ by convention.

In this formulation, f(x) refers to the information element measure (or model facet rating) and $\mu(A_i)$ is the importance of element subset A_i which is the set of elements where $f(x) \geq f(x_i)$. The above expression can also be rearranged to the following form [37, p.143] which is a weighted aggregation of values using the *marginal* increase in element subset weights:

$$C = \sum_{i=1}^{n} (\mu(A_i) - \mu(A_{i-1}))\ f(x_i) \qquad (3)$$

where $\mu(A_0) = 0$, $\mu(A_n) = 1$ and $\mu(A_1) \leq \mu(A_2) \leq \ldots \leq 1$;
and A_i is the set of elements where $f(x) \geq f(x_i)$.

The global Choquet aggregate (C) may be seen as the first moment of f(x) which is why C is sometimes termed the Choquet Expected Value (CEV). Also, by aggregating intervals the Choquet integral can address the significant problem of aggregating measures on mixed ratio and interval scales, as are encountered in many strategic decision analysis models, because that is an admissible operation for interval scale measures. The Choquet integral has also been applied to modelling non-linear attitudes to risk [24]. More aggregation difficulties arising from measurement scale limitations will also be discussed further in Section 6.3.

5.3 Section Summary

In this report the term "aggregation" is taken to mean the induction of global measures. A variety of computational approaches to unconstrained information synthesis has been summarised, highlighting the difference between additive and the non-additive aggregation techniques that are required when there is some form interdependency present in a model. Some axioms were also proposed as the minimum fitness requirements to guide the selection of adequate operators and computational methods. Finally, the Choquet integral was introduced as an example of a non-additive operator which is a function of sets of elements, rather then being a function of individual measures as most operators are. It is suggested that this feature enables the Choquet integral to capture various non-linearities which are present when interdependencies exist between model elements. Another benefit of the Choquet integral is that by aggregating differences it is an admissible operator for interval scale measures which should not be individually aggregated.

6. Uncertainty in Decision Theoretics

6.1 Problem Definition

This section will describe various types of U that may be introduced into the process of modelling complex decisions by means of a particular decision theoretic. Uncertainty in this context refers to fundamental information limitations on the global measures computed, as determined by the adequacy of the computational methods and model fit to the real-world problem characteristics. Thus, rough or over-simplified models naturally yield computational results with less meaning compared to models that reflect the real-world characteristics more faithfully. The notion of Levels of U as introduced in Section 2 will be used to describe the different types of fundamental U that may be introduced in a decision theoretic. Most of these U levels fall into the class of structural U that is the focus of this report. A multitude of decision theoretics have been proposed, and *for convenience,* some of these will be divided into different schools such as multi-criteria decision analysis, multi-attribute utility methods, the European school, and the Russian school. (It should be recognised that these are somewhat loose categories, and the first two are often considered to be the same.) However, neither the complexity of a technique, nor the approach of any single school, enables a method to be totally immune to the following theoretical difficulties and accompanying uncertainties. The intent of this section is not to explain the details of the many methods that have been proposed, but rather to highlight some very basic issues that can limit the value of modelling results.

Decision analysis methods have been divided [58] into two main categories: normative methods and descriptive methods. Normative methods produce advice for the decision maker based on a theoretical model, while descriptive approaches attempt to capture behavioural and contextual effects, perhaps with the aid of experimentation. French [34] has suggested that the present situation differs from the above dichotomy because many sophisticated models contain elements of both approaches. With this understanding, the schools that will subsequently be described differ in the degrees to which they attempt to capture cognitive aspects and contextual relationships. Howard [40] has also suggested that there are three key aspects of a decision problem that help to determine the suitability of a particular technique:

- the number of factors to consider
- the importance of the time domain
- the uncertainties present (types and degrees).

While these aspects may seem to be fairly obvious, the implication is that careful thought should be given to each of them at the start of an analysis. Various authors [14,21,52,57,60,61,63,81,87,90] have also attempted to compare the behaviour of different methods through tests, usually comparing implementation and decision-maker comprehension aspects, or consistency of results for diverse sets of simulated element ratings. Although some of those results concerning practical aspects are justifiable, conclusions about the accuracy of a method's results can seldom be reached due to the difficulty of validating any results, or finding "true" optimal decisions for numerical induction models in general. Also, care should be exercised when making inferences about a method's rigour on the basis of result *consistency,* since they may be consistently bad. Another characteristic related to consistency is "rank reversal", which means that the relative preference order of alternatives may be changed when a new alternative is added to the set of alternatives. Although there is an ongoing debate [77-80,85] about whether this feature necessarily debilitates a method, the significance of rank reversal is largely determined by the features of a problem, and how rank reversal is induced. In some cases it may represent non-linear human cognitive processes and thus be acceptable. In other cases, it may be an anomaly due to the details of a method and therefore it would add unwanted Level 3 U to the results.

Generally speaking, the field under study is referred to as decision making under uncertainty, whereby uncertain information representing evaluations of different aspects of a complex problem is combined to make a global evaluation or choice. With the exclusion of probabilistic inferencing which is not the prime focus of this report, sometimes called decision making under risk, to date the majority of decision theory has addressed the elicitation and aggregation of information relating to human preferences. Characteristically, a mathematically oriented approach derived from traditional OR methods is applied to maximise mathematical functions of decision variables. Typical concepts used in this approach are subjective expected utility, marginal rate of return, partial derivatives of decision variables, the optimal decision space, multiple objective functions, and factor interdependency functions. However, such mathematical techniques frequently require substantive assumptions that can also mask some fundamental structural problems. In recent times, the adequacy of some popular methods has been questioned by several authors [12, 30, 80, 81]. Furthermore, even after 50 years of decision theory development, the important question of tradeoff rationalisation has not really been adequately answered by this function-based mathematical approach to decision analysis. Many solutions are based on exchange ratios [38] which are still subject to some structural difficulties arising from interdependencies. With the previous notions of French, a prescriptive approach to decision modelling combining descriptive analysis with normative methods, would seem to be the most prudent approach in general. This has been called a “syncretic” approach [96] where it is important to first understand the behavioural and information complexities existing in a problem (the descriptive part), before proposing a mathematical technique to address these features.

Fundamentally, decision analysis techniques combine information elements which are evaluations or ratings associated with the various decomposed facets of a problem. Weights may also be applied to indicate such things as importance levels, reliabilities, or credibilities. The main stages of decision analysis are: problem definition by decomposition, facet evaluation, and finally information synthesis. Most decision models then apply hierarchical weighted aggregation using additive summation, or else, additive distance measures based on the Euclidean space assumption. Generally speaking, the problems that are suitable for this decompositional approach are high-level strategic problems. On the other hand, lower-level dynamic and interactive decision problems, as exemplified by military tactical decision problems which are highly time critical, do not in general fit the “divide and conquer” approach. A method which may be more suitable for that type of problem is the Naturalistic or Recognition-Primed Decision Making technique [48] which searches for a historical decision in a human knowledge-base of experience that approximates aspects of the current problem. Case-Based Reasoning is a branch of artificial intelligence that attempts to automate that approach. In contrast, the numerical induction approach described in this report for strategic decision making, initially requires careful consideration of the information inputs and the implicit U forms present to maximise the meaning and value of the computed global metrics.

This section will highlight the uncertainty forms and model misfits that can be found within some common approaches to complex decision analysis. But before describing some of these modelling difficulties, the manner by which decomposed decision information is *rated* before synthesis will be reviewed. This is an important consideration since the nature of the problem determines which rating methods are feasible, and certain options may also introduce structural uncertainties into the computations.

6.2 Decision Element Evaluation Methods

A complex decision is based upon the synthesis of component information related to the facets of the problem that have been defined to *bound* the description of the problem. The definition of the facets is a problem in itself, and may introduce Level 1 U due to a lack of consensus as to what the problem is, as well as Level 2 U as a lack of consensus on the set of facets that need be considered. It is assumed in this section that there is a consensus on the use of a numerical induction model so that Level 2 U relating to the choice of an analytic formalism is absent. Keeney [46] has identified the fact that different decompositions of the same problem can be derived from different viewpoints, or initial assumptions. He points out that if one starts with a range of alternative solutions, called Alternative-Based Thinking (ABT), a set of facets are derived which are highly dependent on the alternative set choice. A recent study [54] has also attested to this. On the other hand, if the problem is viewed in the light of some fundamental values that must be addressed by the solution, a different although probably overlapping set of facets may be derived. Keeney's Value-Focused Thinking (VFT) alerts us to the fact that ABT has historically been the dominant decomposition mode, although decisions are inexorably determined by the broad value system that underlies the problem. The following classification of evaluation methods will neglect this decision decomposition problem, and will assume that a meaningful and adequate set of facets, or problem attributes, has been initially agreed upon somehow.

Facet values (x) and their respective importance weights (w) must be initially evaluated before the aggregation stage that develops global decision values. Figure 9 illustrates the possible ways that these fundamental elements of decision information may be evaluated and some of the potential dangers. These methods for evaluating measures for the decision elements are partially determined by the inherent nature of each facet, and it is important that the concepts associated with facets have a clear meaning. We will commence by defining two main categories of decision problem: the first category being selecting between alternatives, and the second category being the evaluation of a complex situation (i.e. only one alternative). Decision theoretics have evolved primarily to address the first category of evaluative decision analysis, in which facet ratings represent the *facet value* with respect to a decision maker's *value* system. Consequently, how to model cognitive effects in human preference information has been a major thrust with this type of decision analysis. Also important is how to model the U forms in subjective ratings especially when they are multiply aggregated. For example, the importance of distinguishing between ambiguity and randomness in information elements has been highlighted [17,93].

Although both selection between alternatives and the evaluation of a single complex model may require human subjective evaluations, the evaluation of complex situations with diverse heterogeneous facets may include more objective behavioural measures as factor performance measures. These may be quantitative with units, or be normalised qualitative measures without units. The different types of input information between the two decision categories also gives rise to some different structural problems. For example, the meaning of *weights* in the model may be different: one being the relative importance of a facet, and the other being a value exchange rate as how much of one variable is equivalent in value to a unit of another.

Rather than a tradeoff measure, a relative importance weight is a multiplier based on a proximity relationship that indicates the significance of one facet of a problem in relation to the higher-level concept in a model. The meaning of the facet measures themselves may also be different between the two types of application. For example, if the single complex problem being evaluated is military force “Reachback”, the diverse facet measures could be interpreted as performance measures (or degrees of achievement) in relation to predefined standards. This would be quite distinct to a subjective evaluation derived from a decision maker. Of course, both subjective and objective measures can coexist in a single model and need to be combined into a global measure. Depending on the nature of the facets, quantitative or qualitative measures (or both) may be appropriate for the ratings, and several types of rating method for evaluating decision elements will now be described. Figure 9 summarises these methods and the potential dangers associated with each of them.

Quantitative Facet Variables

When facets have a quantitative definition, ratings are usually quantitative measures of their properties with units. In these cases, objective behaviour rather than human opinion is usually the basis of measurement. The measures for different facets may be of similar units (homogeneous) or different units (non-homogeneous). A common variable with homogeneous units is "Cost". However, "Cost" may also have non-homogeneous units such as expected number of casualties in an operation or days to complete a mission.

Aggregation of quantitative information with *homogeneous* units can be termed simple accounting aggregation. Consider, for example, the evaluation of the production revenue of a mining company division with gold and silver mines in different countries which have different tax requirements: X_I = weight produced, and C_I = value per weight. Then two levels of aggregation determine the revenue as the sum of the production weight in each mine by the international market value of unit weight by the percentage retained earnings after tax. With such accounting type aggregation the units of the *quantitative* weights (C_I) must match the facet measures i.e. value silver (or gold) per unit weight and weight of silver (or gold). In this case the ratio of the importance weights (C_I) takes the meaning of exchange ratios i.e. ratio of gold and silver unit values.

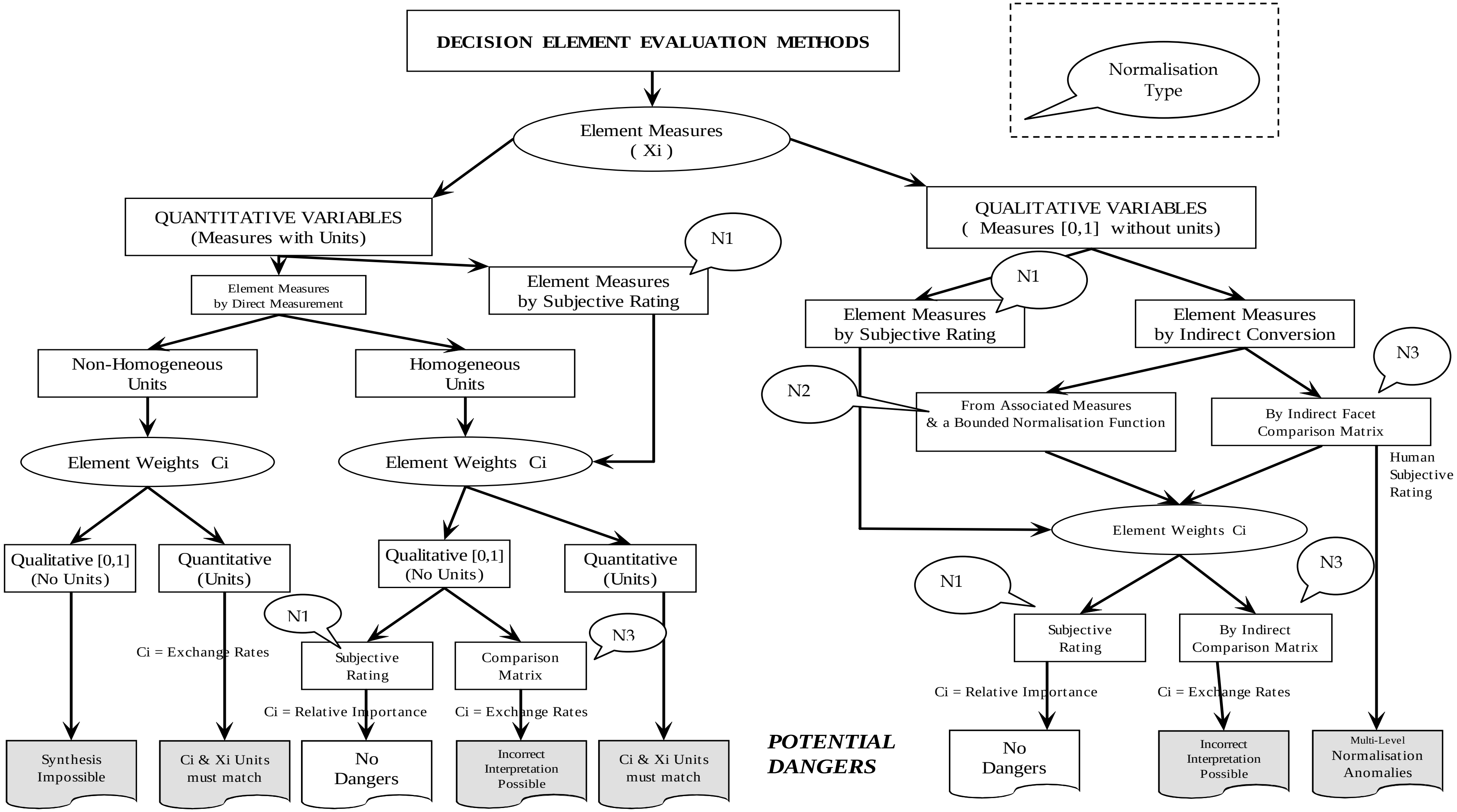

***Figure 9**:* *Potential Dangers in Determining Measures for Model Elements*

Alternatively, quantitative measures with homogeneous units may also be aggregated by using *qualitative weights* on the unit interval [0,1] without units, to represent relative importance (such as low, medium, high) and not be exchange ratios. For example, when facets represent the profit in different countries (and not the production weights) the qualitative weights may be linguistic measures to represent the retained earnings proportion after tax. In this case, the weights would be directly assigned for the different taxation rates, but for other problems the relative weights may need to be subjectively assigned by a decision maker using a facet comparison matrix. While there are no structural dangers in aggregation using directly assigned subjective measures for weights, there are potential dangers when using comparative evaluation techniques, especially with multi-level aggregation. This problem will be explained in more detail in Section 6.3.4.

However, the *magnitude* of quantitative elements with homogeneous units, measurable by definition, may also be subjectively rated when they cannot be measured for some reason. For example, a five point linguistic scale (Low, MediumLow, Medium, MediumHigh, High) may be converted to numerical measures within [0,1]. In this case, based on their units, they could be aggregated with weights derived either quantitatively or qualitatively as shown in Figure 9.

Finally, aggregation of quantitative measures with *non-homogeneous* units simply requires the weighting constants to be in units that convert the facet measures to a common unit. For example:
If the Total project cost =construction cost+over-run penalty cost+workers injury cost. Then # month over-run by $ penalty per month and similar injury conversion to costs, reduces all measure to $ cost. Thus, with non-homogeneous quantitative facet measures care must be taken to match the weight constant units to the facet measures.

For quantitative measures with non-homogeneous units, qualitative weights [0,1] should not be used to aggregate measures.

Qualitative Facet Variables
Certain models may require variables with an inherent qualitative meaning to be evaluated using measures [0,1] that describe positions on the unit interval relevant to some standards. Such measures reflect proximity distances on an interval scale, rather than quantifying a property (as on a ratio scale). Moreover, rating measures for qualitative facets need not be restricted to [0,1] and may also be fuzzy or approximate estimates. Measures for qualitative variables can be derived by human evaluation in three ways: by direct subjective assignment of measures, be derived indirectly by pairwise facet comparisons using some matrix evaluation technique, or by normalisation using a bounded value function. Related normalisation methods (N) will subsequently be described in more detail in Section 6.3.2.

Then the importance *weights* for qualitative variables must be dimensionless numbers, and these may again be evaluated by human judgment directly as relative importance weights, or indirectly as exchange ratios by pairwise comparison ratings across facets (as depicted in Figure 9). Whenever multiplicative operations in matrix evaluation methods are applied, weights as exchange ratios are implied because the ratio scale interpretation with an absolute zero must be adopted to allow multiplicative operations according to measurement scale constraints.

The purpose of detailing the different ways that facets of a decision model may be rated, and their importance weights evaluated, is to identify how uncertainty and anomalies may be unwittingly introduced by these rating methods. The various dangers or pitfalls associated with the different evaluation methods are itemised at the base of Figure 9.

6.3 Some Theoretical Difficulties

The theoretical implications of these different methods for rating facets and weights will now be described, plus some fundamental issues that can affect the adequacy of computations in a decision theoretic. These problems are implicit in many decision techniques and further descriptions of some of these dangers can be found in [12,66,85,97].

6.3.1 The Meaning of Weights

Level 1 uncertainty concerns the lack of agreement on the problem under study. Usually this refers to a large scale complex problem with many diverse and diffuse aspects as viewed by stakeholders with different experience, biases, and needs. However, this level of uncertainty can also be associated with the evaluation of decision factor weights. When decision analysis is considered as the weighted synthesis of elemental information, in this process the meaning of elemental weights is crucial since it can determine the validity of a computational method. There can be four (at least) interpretations of weights as follows:

1. Exchange ratios between elements i.e. the quantitative worth of one unit of an element in terms of another (dependent on the units)
2. Criteria satisfaction priorities.
3. Relative importance values i.e. the significance of the information element or concept. This may be associated with the reliability, credibility, or validity of the information sources.
4. Significance of a decision facet (factor) based on the divergence of that facet's ratings across alternatives; i.e. if all alternatives have similar ratings for that facet then it contributes little information for discriminating between alternatives.

Interpretations 2 and 4 are less common than the other two. One approach to addressing interpretation 2 can be found in [47], while interpretation 4 may be implemented using the information entropy concept [43], or the degree of correlation between columns of the decision matrix [28]. Nevertheless, any of these interpretations, or their combinations, may appear in a decision model and one needs to consider how they should be reflected in computations. This point is discussed also by Barzilai [12].

For quantitative variables with different units (as in Figure 9), weights in an aggregation function would be of the exchange rate meaning to convert the respective units to a common unit. However, for dimensionless qualitative variables without units, perhaps on the interval scale [0,1], the meaning of weights as "relative importance" is not so clear. Weights in this case could simply describe the *significance* of the elemental valuations based on the credibility or reliability of their source. This interpretation imputes a degree (%) of belief meaning on a ratio scale (without any comparative meaning).

However, a pairwise relative importance scale, for example the usual AHP scale {1,3,5,7,9}, can have three meanings: quantitative exchange rates, qualitative relative distances on an interval scale, or simply represent order. Which meaning is appropriate depends on the situation and it determines the admissible mathematical operations in the matrix evaluation method. If a technique uses multiplication or division then the relative importance measures should be ratio scale measures with an absolute zero (i.e. 9 = 9 times greater as for exchange rates, rather than 9 = extremely more important). In other words, the mathematical operations applied to matrix elements should be driven by the type of meaning a rater can embed in a measure. Barzilai [8] discusses the problem of deriving meaningful weights from a relative comparison matrix. Several techniques [7,68] have been proposed to address this particular problem, usually by comparing differences of facet measures as required for interval scales. However, the use of direct subjectively assigned individual weights on an interval scale (N1 in Figure 9) can avoid these computational limitations. Because of the problems associated with the meaning of weights and how to determine their values, some authors suggest that the use of weights should be avoided altogether. Two studies [26,35] that do attempt to compare multi-attribute weight measurement methods are largely inconclusive, concluding only that obtaining consistent estimates may be a function of the number of decision factors the decision maker has to simultaneously compare. Comprehensive discussions of many problems and dangers concerning the use of weights can be found in [5,66]. The main point is that an analyst should be aware of these dangers and the computational limitations of some techniques.

6.3.2 Normalisation Procedures

In order to combine heterogeneous forms of information relating to fundamentally non-commensurate facets, standardisation or normalisation is frequently used to convert dissimilar quantitative or qualitative ratings to dimensionless ratings on the unit interval [0,1]. For normalised measures to be aggregated they need to be on a similar scale with comparable units to comply with the aggregation axioms of Section 5.2.1. A detailed discussion of the problematic aspects associated with normalisation can also be found in [67]. Normalisation is essentially a re-scaling process transposing measures to a decimal interval scale. The sum of normal measures [0,1] need not necessarily be unity, and three types of normal measures (N1, N2, N3) that may be used for evaluating facets in decision analysis are identified in Figure 9 with the following meanings.

N1 = Directly (subjectively) assigned numbers [0,1], e.g. {0.3,0.4,0.8,0.6,0.9}

N2 = Indirectly derived by converting an input quantitative measure via a bounded value or preference function, e.g. conversion of above set where 0.4 = Min (0) and 0.8 = Max (1), with a linear value function yields⇒{ 0, 0, 1, 0.5, 1}.

N3 = Indirectly derived from a mathematical function (such as Euclidean distance) which often requires some integral of the measures to equal unity. This can be called vector normalisation where each measure is divided by the same *constant* (K) to change the scale. N3 as used for comparison matrix synthesis at a *single* level of a decision model, causes no problems since the constant derived from that set of measures has the same effect on each matrix factor and proximity relationships are thus coherent. For any set of measures the choice of the normalisation constant may be determined by the matrix evaluation method (e.g. arithmetic or geometric mean, or an eigenvalue). Alternatively, it may be an arbitrary choice such as the maximum or minimum. Certain normalisation constants are variants of the Generalised Mean Operator (also known as the Minkowski metric) :

Minkowski metric $$K = \sqrt[q]{\sum_{i=1}^{n} (w_i)^q} \tag{6}$$

q = 1: City block measure $$K = \sum_{i=1}^{n} w_i \tag{7}$$

q = 2: Euclidean measure $$K = \sqrt{\sum_{i=1}^{n} (w_i)^2} \tag{8}$$

q = n and 2: $$K = \sqrt[n]{\sum_{i=1}^{n} (w_i)^2} \tag{9}$$

Potential N1 and N2 Normalisation Problems:
Uncertainty may be introduced through N1 and N2 normalisation by means of the *value function* inherent in them, whether implicit or explicit. Thus, a decision analyst should be aware of the effect of the preference value function and ensure that it is appropriate if it can be made explicit.

Potential N3 Normalisation Problems:
The choice of a normalisation *constant* (associated with a matrix evaluation method for example) is another source of uncertainty. Certain decision theoretics are greatly influenced by the normalisation constant so uncertainty can be introduced when there is an arbitrary choice. (The use of Euclidean distance measures in the TOPSIS decision theoretic for example as will be described in Section 6.4.6.) Although the Cartesian space assumption with orthogonal axes is frequently adopted in multi-dimensional scaling, it could be asked why the facets of a military problem should be implicitly related in this way. Furthermore, the notion of a variable's (or a dimension's) separability [15,27] as applied in psychometric testing where most of multidimensional scaling theory was developed, is not necessarily transferable to military problems. Thus, there are many open questions [5,10,11,15,66] regarding the use of arbitrary N3 constants for normalisation, as well as the use of multi-dimensional scaling measures that are arbitrary in the sense that they are not determined by the problem [5]. It could also be argued from a descriptive modelling viewpoint, that it is better to use a method which avoids such measures if they cannot be justified. A common problem with N3 normalisation is the summation of distance values (measures) that have been derived using different arbitrary constants (such as City block and Euclidean). Such aggregation is not adequate because the proximity relationships are effectively on scales with different units within [0,1] (discussed further in next section).

Another more subtle problem (or uncertainty) that can be introduced by N3 normalisation occurs in hierarchical decision analysis models. Barzilai [10,11,13] has identified how multi-level weighted aggregation of utility or preference measures derived by N3 normalisation, can induce aggregation anomalies as well as the rank reversal problem. The problem is that preference measures are combined which are based on scales with different units due to their different normalisation constants derived from the different sets of measures at different hierarchical levels. By violating one of the aggregation axioms, global measures would then have depleted meaning. And although rank reversal can be explained by psychological factors in certain cases, when it is caused by a technical idiosyncrasy such as this it is hard to justify. More detailed analyses of the hierarchical aggregation problem can be found in [9-11].

6.3.3 Measurement Scale Limitations

The definition of a complex problem facet or attribute, together with the measurement scale used to quantify it, may intrinsically restrict the admissible range of mathematical operations allowed. The true nature of a scale being used may also be somewhat obscured by the use of qualitative evaluations or normalisation procedures. Mistaking an interval scale for a ratio scale is not uncommon. When operations are performed on numbers that violate the inherent limitations of that type of numerical scale an unnecessary form of U is introduced into the numerical induction process. The foundations of Measure Theory and measurement scale constraints are described in [9,15,70,84].

It is generally considered that there are four types of measurement scale: nominal (or categorical), ordinal, interval, and ratio scales. Nominal scales are not numerical but the other three are. Nominal scales simply define classes (crisp or fuzzy with overlap) to which information elements are assigned. For example, the weather is fine, cloudy or stormy. Nominal scales label objects (individual or group) and the only operation that can be performed on them is the substitution of labels. However, belief measures may also be attached to the assignments on nominal scales. Ordinal scales are used to describe the relative order (rank) of objects (>,<,=). They may be used to define complete crisp orderings or only partial orderings. Ordinal scales are frequently used in decision analysis and the full range of arithmetic operations are not admissible for ordinal scale measures. They are only invariant under mathematical transforms that are monotonic and order preserving.

The third type of scale is an interval scale with an arbitrary zero point, where it is the difference between measures that is of significance. Interval scales embed information on rank order and equality of intervals, and are invariant under any affine transform, $y = ax+b$, of the true magnitude x with two arbitrary degrees of freedom 'a' and 'b'. The two degrees of freedom for interval scales thus relate to the zero point, which is not an absolute zero but a relative datum point (b) from which to measure proximity distances, and the unit size (a). The Fahrenheit temperature scale is often used to illustrate an interval scale where it is meaningful to say today is 7.5 F hotter than yesterday. Although is sometimes said that all arithmetic operations can be applied to the intervals between interval scale measures, there are in fact some limitations on what operations can be applied to intervals between interval scale measures. If the interval scale measures refer to the same concept, and use the same scale unit, division of intervals is admissible. This has been identified [9] as the defining operation for an interval scale. While subtraction of one interval measure from another is also admissible, addition and multiplication of individual interval scale measures is in general not admissible. However, they can be averaged because the arbitrary zero point cancels out, as shown below.

For example, the simple average (SAV) of three interval scale measures 45, 30 and 60, all with a common reference point or datum:

$$
\begin{aligned}
\text{SAV} &= \text{datum (D)} + \text{average interval} \\
&= \text{D} + \frac{(45\text{-D}) + (30\text{-D}) + (60\text{-D})}{3} \\
&= \frac{3\text{D} + (45+30+60)\text{-3D}}{3} \\
&= \frac{45+30+60}{3} \\
&= 45
\end{aligned}
$$

Averaging may be either a compensatory weighted average or some non-compensatory technique (such as the Choquet Expected Value to model interdependencies when they are present). Interval scale measures are frequently encountered in decision analysis techniques as measures of utility or decision facet value. If the measures are based on a normalised scale [0,1] with a common understanding for the zero point, then the measures themselves represent intervals relative to the datum and may be summed. However, there is also some discussion in the literature [9,84] as to whether such a commonly agreed upon zero point can actually exist, as is implicit in the widely used method of summing subjective scores. Since averaging interval measures is admissible, because it is not dependent on the zero point, averaging of one type or other may be a more robust way to aggregate interval measures.

The fourth type of scale is a ratio scale that has an absolute zero. Ratio scales are of the form y = ax, with one degree of freedom 'a' which determines the unit size. Physical properties such as distance and weight usually have natural zero points of the property. This enables the full range of arithmetic operations to be applied to such measures for determining area, volume, and so on. Thus ratios of ratio scale measures have a precise meaning and are unique. They can be interpreted as ratios of magnitudes of a property and are the most informative type of scale embedding information on rank order, interval equality, and equality of ratios.

The most common example of scale misuse in decision analysis is to perform inadmissible *multiplication* when :

- Measures are on ordinal scales.
- Measures are on interval scales (often subjective ratings of a qualitative variable).

Another technique that may also result in inadmissible operations is when linguistic or semantic scales are transformed to numerical measures. As numerical scales may be of different types, so may linguistic scales be also, according to the meaning associated with them. The particular scale type is then preserved on the numerical transformation. If only order is implied in the linguistic measures then the transposed numerical scale is ordinal with the previous restrictions on operations. An example of such a transposed ordinal scale for relative importance are in the AHP decision theoretic where {nil, low, medium, large, extremely large} is transposed to {1, 3, 5, 7, 9}. This scale transform yields ordinal numerical measures because a rating of 9 does not equal 3 times a rating of 3, nor do the intervals between the numbers have any meaning. Only Max and Min operations and ordinal sorting should be performed on these numbers with no other arithmetic operations admissible.

When the property a linguistic scale measures is a ratio scale concept with an absolute zero (e.g. length, probability, proportion), the numerical scale transform is also a ratio scale. For example, the "relative importance" concept in decision analysis may represent the degree (or proportion) of the factor itself that contributes to the aggregate of all factors. While there has been considerable discussion [13,82] about what relative importance means, if the meaning is taken to be the fraction of a factor that is aggregated into a global value due to its significance or proportional influence in the global picture, then it can be seen to be a ratio scale measure for which multiplication with interval scale preference measures is admissible. However, it should be noted that this is not the meaning of "relative importance" as used in the AHP, nor are they tradeoff weights in that technique. This is discussed in more detail in [92].

When the numerical measures are normal [0,1] having been derived from some normalisation process with upper and lower bounds, they are by definition interval scale measures. For such normal measures the interval distances have meaning and the measures are subject to the computational limitations previously described. Normalised measures are frequently used for measuring decision maker preference or attribute utility, with values assigned in relation to the value system of the decision maker. And as previously noted, there are various questions concerning the commensurability of different interval scale measures which may affect the validity of their summation. Barzilai [9,11] has highlighted this problem and has proposed a new approach [7] for developing preference measures that are not affected by scale units. The aggregation of these measures across *all independent* decision factors can then be achieved because the problem of scale units is avoided. However, when there are interdependencies between component factors in a model, or between Barzilai's relative preference measures at the model leaf nodes, some non-additive procedure is really required to synthesise those unit free measures into a global measure for comparing alternatives. As previously stated, the Choquet fuzzy integral is one such non-additive technique that may be appropriate. Furthermore, since individual interval scale measures without an absolute zero should not be summed, it is especially suitable for interval scale measures because it sums and averages measure *intervals*.

In the previous section it was also described how normalisation procedures may result in the combination of measures with different scale *units* when:

- Ratio scale measures for quantitative variables (for example, $ Cost) are N3 normalised using different normalisation constants.
- Normalised measures derived by different methods are combined. For example, N3 normalisation constants with a N2 bounded value function.

A further example of scale misuse commonly occurs in Cost/Benefit/Risk analysis using normalised interval scale measures in division operations to form ratio measures. Alternatively, a non-additive technique may be a more appropriate method to synthesise these three normalised interval measures into a global measure for alternative comparisons because indeterminate interdependencies are usually present between costs, benefits, and risks.

Thus, in addition to the Level 6 U that may be intrinsic to a scale, Level 3 U concerning an inappropriate and inadmissible mathematical operation may also be unwittingly introduced. Also, several studies [55,56,61] have been made comparing the effect on the AHP method of using different measurement scale units. Unfortunately, these AHP scale studies assume that the input comparison matrix ratings are on a ratio-type scale, which they are not. For this reason, these results are of questionable merit. For example, 9 on Saaty's scale means that the numerator element is absolutely preferred to the denominator element, and does not mean that the numerator element is 9 times better than the denominator element. This problem has also been discussed by Stewart [85]. For these reasons, decision analysis computations should always be cognisant of the inherent limitations of the scales being used for the quantification of elemental variables or facets of the problem under study. Barzilai [9] discusses in detail how measurement fundamentals can affect preference evaluation, and other discussions of Measure Theory fundamentals including homomorphic relations and existence of uniqueness theorems can be found in [9,15,70,84]. As discussed in the introduction of this report, no output validation is possible for numerical induction models and no bells ring when operations are applied that exceed the information limitations of the inputs. To ensure that the output embeds the most information and meaning for the given inputs, the best the analyst can do is to accord with the computational constraints.

6.3.4 Comparison Matrix Evaluation

When a matrix of pairwise comparisons of decision facets by an expert is used in a decision theoretic, a number of techniques have been proposed to synthesise the matrix into a "priority" vector which establishes proximity measures on the unit interval [0,1]. (In this section we will ignore the previous scale limitation problem.) These pairwise ratings may be for determining facet weights or facet/alternative decision values or preferences. The range of techniques available for matrix synthesis includes the right eigenvalue, geometric mean (row or column), harmonic mean (left eigenvalue), least squares, constant-sum, and simple row average. Although many comparative studies exist [8,26,27,28,35,61,63,68,78,81,104] there is little consensus as to the adequacy of those methods. Perhaps the most widely used method is that of the AHP right eigenvalue method. Saaty and Vargas [77] have claimed that their basis for selecting this technique is that it "preserves rank strongly" in the presence of rating inconsistencies. However, Barzilai [8] has also shown numerically that the left eigenvalue has exactly the same properties as the right eigenvalue thus yielding a different priority vector. Furthermore, it has also been demonstrated [8,35] that the geometric mean better satisfies fundamental consistency requirements for multiplicative matrices, while the arithmetic mean does so for pairwise additive matrices. When using the AHP for group decision making, Zhou [104] has compared various AHP variants and concluded that the technique for comparison matrix evaluation has less effect on the decision than the method for aggregating the ratings. Unfortunately, many of these results are compromised by the inherent problem of different normalisation constants in the hierarchical aggregation of preference values.

In this way, U can be introduced into computations through an inadequate matrix evaluation method, and it is Level 4 U concerning computational micro-mechanisms in the U framework of this report. In order to minimise this kind of U, some reasonable justification is desirable for the adoption of a particular method for evaluating the matrix of comparative ratings, in relation to the type of scale used for the ratings. Overall, there is some doubt as to the validity of many matrix evaluation methods because the measures are interpreted as ratio scale measures when the meaning of the decision factors and method of measuring induces interval scales.

6.3.5 Decision Element Interdependencies

As discussed in Section 4.3, interdependencies of several types may be present in a decision problem. When element ratings represent human preference evaluations, they are preference interdependencies. But if the element ratings are objective behavioural measures, or payoffs, they are generally called utility dependencies. For example, elements within a cost hierarchy of a complex project can also influence the ratings of the benefit and risk hierarchy elements. This type of interdependency is usually of an indeterminate degree because the cost elements do not totally determine the benefit and risk elements, since various design and intrinsic characteristics also contribute effects. Nevertheless, conventional Cost/Benefit/Risk analysis often combines the three individual global values using additive operators as if the hierarchies were independent. Usually such relationships are qualitative in nature and cannot be accurately quantified nor expressed in mathematical expressions. But besides facet value interdependencies, facet importance weights may also be subject to interdependencies. When they are subjectively determined they are preference dependencies. But they may also be due to some non-preferential endogenous characteristics such as reliabilities or credibilities. In general, interdependencies between importance weights are also difficult to quantify since they are usually assigned qualitative ratings in strategic decision problems. Dependencies of any kind may also exist between only a small subset of facets within a model. In these cases a multi-tier model (as in Figure 6) may enable these links, which may cross different levels of abstraction, to be clearly identified so that they can be addressed appropriately in computational methods. Another approach to representing interdependencies is to use a planar graphic network model. Saaty [76] adopts such a model in his recent Analytic Network Process (ANP) for decision making with interdependencies. However, like the AHP, the ANP is also debilitated by measurement scale issues.

The following example will be used to illustrate how different levels of interdependency may be present. Consider, the computation of a global performance measure to monitor the state of a military campaign. For simplicity, let this evaluative model have four main components: Blue Cost, Blue Gains, Blue Losses, and Blue Combatant State. For simplicity it is assumed that no Red state information is available. The integration of these four major aspects can be viewed as a complex bubble representing the interactions between these four variables. Also, while global Costs may be evaluated using an accounting type tree hierarchy, interdependencies would exist within the other three variables. For example, if the Blue Combatant State is evaluated from the ratings of ten behavioural characteristics, the observed behavioural facets may be conditionally independent of each other but they may be associated via the meta-variable of group morale. So the method of synthesising the ten different behavioural ratings, which may themselves be heterogeneous types of measures, should capture such associative dependencies in the global Combatant State value. Also, the amount of Blue Gain may be related to the amount of Blue Loss, so the overall campaign state value should be based on a procedure that is cognisant of the interdependencies between Gains, Losses, Combatant State, and Costs.

In general, the conventional mathematical approach to resolving interdependencies in numerical induction requires the development of expressions or objective equations which require explicit dependency relationships. Carlsson and Fuller [22] adopt this approach (see Figure 8). Multi-objective programming, such as goal or compromise programming, is also used to capture the tradeoffs caused by objective interdependencies. However, such mathematical approaches, including the connectionist ANP technique, cannot address the vague interdependency forms that usually exist in data-sparse and abstract strategic decision problems. All forms of interdependency effectively introduce non-linearity into the process of information aggregation, and thus require non-additive and non-compensatory techniques. Uncertainty that may be introduced into global decision variables when an inadequate dependency modelling process is used is termed Level 3 U in the framework of this report. That level refers to the adequacy of the general computational macrostructure with respect to invalid aggregation procedures or methods that do not fit the characteristics of the problem.

6.3.6 Tradeoff Rationalisation

Tradeoffs between multiple objectives are a common feature of complex problems which need to be rationalised in information synthesis. When a mathematical formulation of a problem is possible with objective functions and constraints, optimal decision theory can be applied to determine values of variables that yield optimal payoffs. But information sparse discrete decision analysis problems should be treated in a non-optimal manner, more like Simon's "satisficing" concept to find good-enough choices which satisfy minimal aspiration levels, or else the best of a set of alternatives (Pareto optimal). How to rationalise tradeoffs in decision analysis techniques has been a vexing question for many years and a recent treatise [38] by some seminal thinkers in the field (Hammond, Keeney, and Raiffa) illustrates the difficulty of the problem.

The approach recommended in that recent treatise progressively deletes dominated solutions by modifying the measures for individual objectives or criteria, one at a time adjusted according their weights (as exchange ratios in this case), until a single non-dominated alternative is forced to appear. This is achieved by first determining the change necessary to cancel out one objective (or criterion) and then by determining what change in another objective (or criterion) would compensate for that change. In this manner, dominated alternatives are eliminated until only one remains which is taken to be the best solution. The flaw in this approach is that with *systemic* inter-related information, individual measures cannot be separately adjusted because a change in one measure can result in unknown changes elsewhere in the system of measures due to interdependencies. In other words, this approach only works for strictly independent sets of criteria ratings. And since strategic decision problems are invariably described by sets of systemic information with associative, if not causal interdependencies, the approach is inappropriate.

Adopting another approach, if the multiple objectives and their components are considered as a system, measures may be integrated upwards factoring in their mutual disparities as previously described. In this manner tradeoffs between global decision variables are rationalised on the basis of their consistency because divergent low values represent windows of fallibility in system models. Elemental weights in such models may be reliabilities, element importances, or information credibilities. Uncertainty introduced by an inadequate tradeoff modelling method can be viewed as Level 3 U, which again refers to the computational macrostructure features.

6.4 Uncertainties within Some Popular Decision Theoretics

The primary objective of this section is to highlight the manner by which the previously described structural uncertainties (the Levels of U) may be tacitly present in some widely used decision analysis methods. Only a brief overview of these methodologies will be used to describe where the different types of U are present, and the details and full descriptions of these decision theoretics can be found in texts such as [1,31,33,40,42,45,62,87,94,100].

6.4.1 Multi-Attribute Utility Theory (MAUT)

This is a major school, dominant in the US, which assumes a decision problem can be modelled by a real valued function that can be maximised between alternatives. Some authors [94] have suggested that MAUT has two variants: the Harvard school of Keeney and Raiffa [45], and the Stanford school of Howard [40,41]. A key difference between them is that in the Stanford school it is not necessary to compute a global measure to make a decision. In 1947 von Neumann and Morganstern [89] axiomatised expected utility theory and thus laid the foundations of MAUT, as applied to econometrics. Accordingly, the expected utility of an alternative course of action was taken to be the weighted average of the component utilities, using additive aggregation where the weights were the probabilities of component outcomes. But over the years, beginning with Allais [2,3] and the Ellsberg phenomenon [32], there has been criticism on the simple weighted aggregation mechanism used to determine the expected utility. (A comprehensive summary of various issues concerning the interpretation and application of MAUT can be found in [31].) In short, there may be problems where non-additive aggregation is more appropriate, either due to the human decision process or preference idiosyncrasies, or due to interdependencies between different facets of the problem. In general, facets of the problem in MAUT are evaluated and converted to a scale that is assumed to have common or commensurate units, sometimes called "utils". However, as discussed in Section 6.3.3, on measurement scales, some doubt exists about the validity of this assumption. At other times, Utility functions may be based on probabilistic risk lotteries and synthesised by simple weighted aggregation of values. Defined value functions then convert facet ratings to utility measures. The problem of weight determination is inherent in MAUT, and when weights are derived from comparison matrices uncertainty may be introduced via the matrix evaluation technique, as well as via hierarchical weighted aggregation due to the multi-level normalisation problem. A collection of recent literature pertaining to utility theory can be found in [6] and overall MAUT is vulnerable to Levels 3 and 4 uncertainty.

6.4.2 Multi Criteria Decision Making (MCDM)

MCDM is a collection of techniques that differ somewhat from classical utility decision theory. Zeleny [100-103] was a key figure in the development of MCDM and placed emphasis on how to match mathematical methods to the human cognitive process. However, the term "MCDM" is on occasion also used to describe the whole field of "divide and conquer" decision modelling, and to describe any method using rating or scoring of a list of criteria. At other times MCDM is used to refer to relative preference modelling through comparison matrices. So MCDM may be somewhat broader than MAUT, embracing a potentially wider range of mathematical techniques. An early proposal of MCDM was interactive multiple criteria programming, for situations where a decision maker can only evaluate weights or factor ratings when confronted with a dynamic problem. In general MCDM is open to the same dangers as MAUT, perhaps with a greater emphasis on the problems associated with comparison matrix evaluation: the meaning of the comparisons, the use of a scale without an absolute zero, and the matrix evaluation technique. In addition, the danger of multi-level normalisation anomalies is also present. Thus MCDM is also exposed to a broad range of Level 3 and 4 U forms.

6.4.3 The Analytic Hierarchy Process (AHP)

This is a popular version of MCDM (or some would say MAUT) with a user-friendly software implementation called Expert Choice. The method [74,75] is founded on the evaluation of pairwise comparison matrices and unnecessary U is introduced into the matrix evaluation in two ways. Firstly, ordinal input ratings are interpreted as ratio scale measures and subjected to inadmissible multiplication operations in the eigenvector evaluation technique. Secondly, the priorities are determined from the right-hand eigenvector associated with the maximum eigenvalue without any sound justification for doing so (the justification given that it averages out inconsistent ratings and so preserves the true rank order being questionable). Furthermore, anomalies in hierarchical aggregation are introduced by combining normalised measures derived using different normalising constants, and hence being based on scales of different units. A detailed analysis of the various structural problems within AHP can be found in [7-13,30,50,67,78,80,81,88,92]. Overall, very significant combinations of Level 2, 3, and 4 U are present in an AHP application and the method should be used with caution.

6.4.4 The European School of Outranking

The history and key features of this school are described in [72,73]. Many of its methods, as exemplified in ELECTRE [71], try to avoid any precise conclusions in an effort to model the inherent vagueness in subjective ratings, and the impossibility of computing a "true" decision value (or utility) for an alternative. The methods are largely based on partial ordering of preferences (since they are only rough estimates anyway) rather than the crisp dominance concept. Given a set of alternatives and the decision facets, the method reduces the non-dominated set of alternatives using a “concordance” index to measure the relative advantage of an alternative, and a “discordance” index to measure its relative disadvantage. These act as distance measures and an outranking algorithm produces a “kernal” (a subset or shortlist) of alternatives which overall are approximately the same. The decision maker must then decide what extra information is required to select between them. The outranking algorithm is complex and sensitive to arbitrary parameters. Additionally, weights must initially be input by the decision maker so it may be open to the weaknesses of comparative matrix methods. Overall, the complexity of the technique introduces various forms of Level 2, 3, and 4 U in the attempt to capture the vagueness in the decision analysis process.

6.4.5 The Russian School (ZAPROS)

The ZAPROS technique, developed by Larichev and Moshkovich [51,52] uses qualitative evaluations by the decision maker to develop partial orders on possibly large sets of alternatives. The method is said to address the difficulty humans have to express preferences consistently, especially over large sets. To simplify preference elicitation, the decision maker only chooses between two alternatives differing in one criterion only. Preference independence is assumed across facet values and pairwise comparisons are elicited. Like Outranking, the method provides only a partial rank order and does not guarantee a complete rank-ordering. The key notion of this method is the concept of a joint ordinal scale developed from the input preferences of the decision maker from the micro-comparisons based on a predefined categorical scale. The method is relatively robust for non-systemic sets of information, has good transitivity control, and minimises judgment problems with large sets. When applied to systemic information problems with interdependencies, in addition to Levels 3 and 4 U, it is less robust with extra susceptibility to Level 5 U in evidence quality.

6.4.6 TOPSIS

TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) is a method [42] based on the notion that the preferred alternative should have the shortest distance from a hypothetical ideal solution, plus the greatest distance from the worst hypothetical solution, assuming Euclidean space distance measures. The decision matrix is first normalised by dividing by a different constant (similar to root mean square) for each column. This normalised matrix is multiplied by input subjective weights for each column (criteria). The best and worst values are selected from each column to yield the ideal and negative-ideal solutions. Then for each alternative (row) a separation measure is computed from these two vectors assuming Cartesian space. The final ranking is based on a decision metric (the "Closeness" measure) that is a function of both separation measures. The results of TOPSIS are conditioned by the normalisation method and the Cartesian space assumption, which is Level 3 U concerning macro-computational uncertainty. Level 4 U variants are also present due to the *different* normalising constants between columns which means that measures of different units are combined. Furthermore, the manner of incorporating weights is inadequate because they are squared in the distance from ideal evaluation which distorts their intended meaning. Thus, this popular technique has considerable Level 3 and 4 forms of uncertainty.

6.5 Section Summary

This section has discussed some fundamental types of uncertainty that may inadvertently enter decision analysis computations and affect results. The vulnerability of some well-known decision analysis techniques to these uncertainties has also been described in very broad terms. Some of these decision analysis techniques, as exemplified by the AHP, are especially susceptible to a variety of these fundamental types of uncertainty. And unfortunately, the popularity of a particular method or its widespread use does not necessarily indicate the adequacy or validity of the method. Consequently, the output of many decision analysis techniques should be treated with some caution, or at least be cross-validated with other approaches. No method is always the best from the uncertainty viewpoint. Moreover, the need for a keen appraisal of the characteristics of input information to a model has been stressed before selecting a suitable decision analysis technique. Such a preliminary information analysis should also examine the relative merits of different analytical formalisms such as choosing between probabilistic models with a priori conditional probabilities, or numerical induction models with subjective ratings.

Four stages in modelling complex decisions have been identified:

- Identifying the component considerations (facets, factors, or attributes)
- Identifying any inter-relationships that may exist between the components.
- Evaluating and determining performance measures for component factors (preference or utility measures) and their importance weights (if any).
- Aggregating the component evaluations into a global decision measure by numerical induction, or into a global preference order of alternatives by other methods.

Uncertainty may be introduced in the first stage when important component factors are not identified. The second stage may result in a model that may be hierarchical with independent factors, or multi-tier with interdependent factors, or a flat planar network of factors. Uncertainty is introduced at this stage when a problem is inadequately conceptualised and the model does not capture some important inter-relationships existing between components. Uncertainties may be introduced in the last two stages through invalid computational procedures. Very often in numerical induction procedures the fourth stage of aggregating component measures is performed using additive weighted hierarchical aggregation. Apart from the problem of multi-level aggregation of measures that have been derived using different normalising constants (as discussed in Section 6.3.2), hierarchical aggregation is considered to be unsuitable for many strategic decision problems because of ubiquitous interdependencies between strategic type factors.

7. Defence Decision Analysis

The identification of unnecessary uncertainty forms that may be intrinsic in the structure of a decision model and the computational techniques applied, can have relevance to the analysis of some complex Defence problems as will be described below.

7.1 Avoidance of Less Robust Decision Analysis Techniques

The definition of six levels of U in structured decision models and their computational methods can facilitate the identification of potential weaknesses in a decision analysis technique. With a greater awareness of the multiple dimensions of U in a technique less robust methods may then be avoided. Moreover, models which suit the actual characteristics of the problem more closely may then be developed, such as capturing any elemental interdependencies.

7.2 Performance Evaluation of Complex System Models

Several types of military decision analysis can be classed as performance evaluation of complex systems where there are diverse non-homogeneous aspects that are loosely coupled, and the objective is to develop an overall meta-measure for the system behaviour. Three examples of such problems will now be briefly described.

7.2.1 Assessment of Military Operations

The traditional way to evaluate a military operation is to use predefined Measures of Effectiveness (MOEs) for each task or activity in the overall plan. Field reports are then rated against these MOEs to determine whether the objectives of the task have been achieved. Quantitative and qualitative aspects usually need to be considered conjointly such as tangible gains/losses and risk and morale aspects. In some military operations assessment systems lower tactical assessments are then hierarchically aggregated to evaluate higher-level objectives. One problem with using predefined MOEs in this way is that it assumes perfect foresight whereby the totality of associated effects is known. The weakness here is that not all consequent effects can be captured in a relatively small number of MOEs, plus unforeseen negative effects may occur which need to be considered that are completely different to any predefined MOEs. Thus, MOEs alone often may not describe clearly the degree to which objectives are being achieved, unless it is a very clear-cut and simple tangible objective.

To address this weakness of using MOEs in the traditional manner, assessment within the effects-based operations (EBO) concept should take a somewhat broader approach. With this approach tasks are linked to desired effects but the assessment of achievements is based upon the aggregate capture of all effects that can be linked to an implemented task and its objectives. Thus, field reports would capture a broader range of effects in the global domain including negative effects plus effects at different levels of detail and importance. In this way, assessment is based on a broader spectrum of reports which may even have global reach, rather than being limited to military aspects in one particular area of operation.

The problem that then confronts the EBO approach is how to synthesise these diverse assessments, positive and negative, important and not-so-important, into assessments of objectives achievement across different levels, from tactical objectives through to strategic and national objectives. This problem can also be viewed as performance evaluation of a complex system with loosely coupled elements, and numerical induction models could be applied with evidence integrated using an algorithm based on the Choquet integral to address the elemental interdependencies. Currently, investigations are progressing in Command and Control Division to apply this methodology for the assessment of military operations in a prototype tool called Plan Monitor. Further discussion of some the difficulties facing the development of military expert systems for threat assessment based on numerical induction models using elemental assessments, can also be found in [69].

Non-military organisational performance evaluation models, such as the Balanced Scorecard method [23,44], may also involve loose couplings again presenting aggregation difficulties both within and between measure quadrants or categories. There is little doubt that it is beneficial to develop measures for a diverse range of facets of a complex system or organisation, and not to base decisions solely on a limited range of considerations such as costs, profits, or some equivalent military variable such as relative attrition rates. However, improvements in one facet, or group of facets, may be linked with deterioration or low values of other facets. Thus, measures should *only* be evaluated in *isolation* if they are independent. In general, it would be better to base any cognitive interpretation of a set of inter-related systemic measures on an integrated meta-measure. Unfortunately, these complexities are seldom addressed in organisational performance evaluation systems (such as the Balanced Scorecard), and evaluation of category measures in isolation with simple weighted addition for measure synthesis seems to be the de facto standard.

7.2.2 Vulnerability Assessment of Complex Architectures

For national infrastructure protection, assessment of the vulnerability of complex infrastructural architectures is required. Such architectures include supply grids for telecommunications, electricity, gas, water and transportation. This supply network is also composed of physical, human and functional process elements, any of which may present windows of opportunity for attack. Furthermore, these various supply networks may be inter-related and their performance also may be influenced by factors beyond national control such as international conflicts, oil prices, or even catastrophic weather events. All of this complicates the assessment of national infrastructure vulnerabilities.

In the past the reliability or vulnerability of large socio-technical systems has frequently been evaluated by the Fault Tree technique of systems engineering. However, that technique is also limited when there are interdependencies within and between systems. To address vulnerability assessment of very large complex systems, numerical induction models may also be applied to develop meta-measures for the different types of system- the technical, human, and process systems. By applying a multi-tier model to represent the complex system (as in Figure 6), aggregate measures could then be successively developed using the Choquet integral to address the loosely coupled features of the system.

7.2.3 Evaluation of Future Capability Requirements

The planning and analysis of future military capability requirements may also be treated as a complex system evaluation problem. Initially the range of feasible future threats would need to be assessed from the whole field of possibilities. Next the range of potential capabilities that could feasibly be brought on-line would need to be identified. When selecting the set of capabilities to be acquired in relation to current assets, various types of functional interdependencies, synergies and redundancies would also need to be assessed. Then for each feasible capability, a Cost, Benefit and Risk analysis would ideally be performed considering the various strengths and weaknesses. Overall, this is another complex system evaluation problem due to couplings between different aspects, as between future capabilities themselves in relation to the range of feasible threats.

Undoubtedly, however, the rapid change of military system technologies combined with a changing threat spectrum does make analysis for long-term capability planning very challenging.

8. Conclusions

Decisions concerning complex strategic type problems are often made in an *ad hoc* or intuitive manner based solely on the experience of the top-level decision maker. While this may be completely appropriate in some situations, it may not always be advisable to make such unauditable decisions. Some examples of such strategic decisions include complex project evaluations, sensitive facility site selection, tender evaluations, company mergers or takeovers, and Defence capability development decisions. Political considerations may also be of paramount importance and can absolutely determine a decision. One step beyond the intuitive or political type of decision making is to use relatively simple methods to score component factors and payoff expectations. A further step upward in formal rigour is to apply a more complex structured decision model to ensure that all important aspects and viewpoints have been considered, including risks, and also to achieve greater consistency when component evaluations are integrated to enable the overall comparison of decision alternatives.

Structured cognitive models that synthesise component evaluations into global measures for decision making have been termed numerical induction models in this report. It is fundamental that the results of this kind of model cannot generally be validated. Using a framework consisting of six levels of modelling uncertainty, this report has described a range of considerations and pitfalls that can influence the quality and meaningfulness of a structured decision model's output. Various structural issues of numerical induction models have been described, covering both the type of cognitive model selected and the type of information aggregation mechanism applied. In isolation many of the problems associated with the described issues may seem to be unimportant, not particularly significant, and too basic to be concerned about. However, when a number of these considerations combine in a model they can have a significant impact on the quality of a model's output. For this reason, the objective of this report was to make analysts aware of implicit assumptions and computational limitations that can influence the quality and hence the credibility of the outputs of models that are used for strategic decision analysis. A selection of some popular decision analysis techniques has also been presented and a variety of the described structural issues identified in each of them.

Therefore, before selecting or formulating a model, it may pay dividends to take a careful look at the characteristics of the problem and the available information inputs. With a closer model fit, more knowledge may then be extracted by the modelling process. The following words of Pomerol and Barba-Romero [66: p.6] illustrate this viewpoint:

"*Multicriterion decision making can now be considered as a field of activity in which practical application and informatics are dominant. Theoretical research is not of course devoid of interest, but it is now more concerned with giving depth to existing ideas than in innovating. On the other hand, the possibilities of informatics have not yet been wholly explored; we may even say that application of multicriterion methods in professional contexts has only just begun.*"

In this quotation "informatics" can be taken to mean computerised information representation and processing methods, and what is described is a fresh approach to decision analysis (i.e.

"giving depth to existing methods") which aims to minimise unnecessary uncertainty introduced by overly complex computational techniques. Instead, the modelling process should be driven by the information forms that are actually available in relation to any complexities that exist, such as interdependencies.

Although this report has focused specifically on numerical induction models, many of the issues and complications described are also present in other types of conceptual models. Therefore paying heed to the introduction of unnecessary uncertainty is an attitude that may have broad benefits for the Operations Research community in general. In recent times it has also been emphasised that Defence analysis should focus more on benefits in relation to costs and risks, not only for platform replacement and capability development, but also to justify other types of change. It is just for this type of analysis that numerical induction models may be applied, and where the model outputs may possibly be compromised due to uncertainty that is unnecessarily injected into the results.

In conclusion, the following recommendations are offered to assist when selecting a numerical induction technique to minimise the introduction of unnecessary uncertainty through invalid operations and excessive assumptions.

For numerical induction models which elicit preference judgments from experts, as well as determining measures from any objective data that is available, the Preference Function Method of Barzilai is a particularly robust and theoretically sound method for determining the individual preference measures for the leaf node elements of the model. And when the leaf-node factors are preferentially independent, this method is also adequate for *aggregating the individual leaf-node measures* into global meta-measures for comparison of alternatives.

However, when interdependencies exist between leaf-node factors some non-additive approach, such as one that uses the Choquet integral, may be more suitable for aggregating the component leaf node measures as determined by the Preference Function Method into global meta-measures for choosing between alternatives. At present, the application of such a non-additive information aggregation technique is being investigated for use in decision aids to support the planning and monitoring of effects-based operations. Moreover, a wide range of Defence applications potentially exists for analysis by numerical induction models, especially in the general field of capability planning.

9. References

[1] Aliprantis, S and Chakrabarti, S. (1999) *Games and Decision Making*, Oxford University Press: Oxford.

[2] Allais, M. (1953) Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole americaine. *Econometrica*, **21**, pp. 503-546.

[3] Allais, M. and Hagen, O. (1979) *Expected Utility Hypothesis and the Allais Paradox*, Reidel: Dordrecht.

[4] Almond, R. (1995) *Graphical Belief Modeling*, Chapman & Hall: London.

[5] Anderson,N. (1981) *Foundations of Information Integration Theory*, Academic Press: New York.

[6] Barbera, S., Hammond, P. and Seidl, C. (1998) *Handbook of Utility Theory: Volume 1 Principles* , Kluwer Academic: Dordrecht.

[7] Barzilai, J. (1997) A New Methodology for Dealing with Conflicting Engineering Design Criteria, *Proc. 18th Annual Meeting of American Soc. for Eng. Management,* pp. 73-79.

[8] Barzilai, J. (1997) Deriving Weights from Pairwise Comparison Matrices, *J. Operational Research Society,* **48**(12) , pp. 1226-1232.

[9] Barzilai, J. (1998) Measurement Foundations for Preference Function Modelling, *Proc. IEEE Int. Conference on Systems, Man, and Cybernetics*, pp. 4038-4044.

[10] Barzilai, J. (1998) Understanding Hierarchical Processes, *Proc. 19th Annual Meeting of American Soc. for Eng. Management*, pp. 1-6.

[11] Barzilai, J. (1998) On the decomposition of value functions, *Operations Research Letters*, **22**(4-5), pp. 159-170.

[12] Barzilai, J. (1999) Decision Theory Misconceptions and Remedies: Implications for Decision-based Design, *Proc. Third Int. Conf. on Eng. Design and Automation,* Vancouver, pp. 767-773.

[13] Barzilai, J. (2001) Notes on the Analytic Hierarchy Process, *Proc. NSF Design and Manufacturing Research Conference*, Jan. 2001, Tampa Florida, pp. 1-6.

[14] Beroggi, G. (2000) An experimental investigation of preference elicitation methods in policy decision making, *J. Multi-Criteria Decision Making*, **9**, pp. 76-89.

[15] Birnbaum, M. (1998) *Measurement, Judgment, and Decision Making*, Academic Press: New York.

[16] Bloch, I. (1996) Information combination operators for data fusion: A comparative review with classification, *IEEE Trans. Systems, Man, and Cybernetics (A)*, **26**(1), pp 52-67.

[17] Bookstein, A. (2001) Implications of ambiguity for scientometric measurement, *J. American Society for Information Science and Technology*, **52**(1), pp. 74-79.

[18] Bouyssou, D. and Vansnick, J. (1986) Noncompensatory and generalised noncompensatory preference structures, *Theory and Decision*, **21**, pp.251-266.

[19] Bouyssou, D. and Pirlot, M. (1999) Conjoint measurement without additivity and transitivity, in *Advances in Decision Analysis,* N. Meskens and M. Roubens (Eds.), Kluwer Academic: Dordrecht, pp. 13-29.

[20] Breese, J., Goldman, R. and Wellman, M. (1994) Introduction to the special section on knowledge-based construction of probabilistic and decision models, *IEEE Trans. Systems, Man, and Cybernetics,* **24**(11), pp.1577- 1579.

[21] Buede, D. and Maxwell, D. (1995) Rank disagreement: A comparison of multi-criteria methodologies, *J. Multi-Criteria Decision Analysis*, **4**, pp. 1-21.

[22] Carlsson, C. and Fuller, R. (1995) Multiple criteria decision making: The case for interdependence: *Computers and Operations Research,* **22(**3), pp. 251-260.

[23] Chang, R. and Morgan, W. (2000) *Performance Scorecards,* Jossey-Bass: San Franscisco.

[24] Chateauneuf, A. (1994) Modeling attitudes towards uncertainty and risk through the use of Choquet integral, *Annals of Operations Research*, **52**, pp. 3-20.

[25] Choquet, G. (1953) Theory of Capacities, *Annals Institute Fourier*, **5,** pp. 131-295.

[26] Chu, A., Kalaba, R. and Springarn, K. (1979) A comparison of two methods for determining the weights of belonging to fuzzy sets, *J. Optimisation Theory and Applications,* **27**(4), pp.531-538.

[27] Crawford, G. and Williams, C. (1985) A note on the analysis of subjective judgment matrices, *J. Mathematical Psychology,* **29**, pp. 387-405.

[28] Diakoulaki, D., Mavrotas, G. and Papayannakis, L. (1992) *Objective weights of criteria for interfirm comparisons*, Thirty sixth Journees du groupe europeen Aide Multicritere a la Decision, Luxembourg.

[29] Dubois, D and Prade, H. (1984) Criteria aggregation and ranking of alternatives in the framework of fuzzy set theory, *TIMS/Studies in the Management Sciences*, **20**, pp. 209-240.

[30] Dyer, J. (1990) Remarks on the Analytic Hierarchy Process, *Management Science*, **36**(3), pp. 228-230.

[31] Edwards, W. (Ed.) (1992) *Utility Theories, Measurements and Applications*, Kluwer: Dordrecht.

[32] Ellsberg, D. (1961) Risk, ambiguity, and the Savage Axioms, *Quarterly J. Economics,* **75,** pp. 643-669.

[33] Figueira, J., Greco, S., Ehrgott, M. (2005) *Multiple Criteria Decision Analysis: State of the Art Surveys*, Springer: New York.

[34] French, S. (1996) The framing of a statistical decision theory: A decision analytic view, *Bayesian Statistics*, **5**, pp 147-163.

[35] Golany, B. (1993) A multicriteria evaluation of methods for obtaining weights from ratio-scale matrices, *European J, Operational Research* , **69**, pp. 210-220.

[36] Grabisch, M., (1995) The fuzzy integral in multicriteria decision making, *Fuzzy Sets and Systems*, **69**, pp. 279-298.

[37] Grabisch, M., Nguyen, H. and Walker, E. (1995) *Fundamentals of Uncertainty Calculi with Applications to Fuzzy Inference*, Kluwer Academic: Dordrecht

[38] Hammond, J., Keeney, R. and Raiffa, H. (1999) *Smart Choices*, Harvard Press: Cambridge.

[39] Hetherington, J. (2000) Role of theory and experimental design in multivariate analysis and mathematical modeling, in *Handbook of Applied Multivariate Statistics and Mathematical Modeling* (p. 37), H. Tinsley and S Brown (Eds.), Academic Press: San Diego.

[40] Howard, R. and Matheson, J. (1984) *The Principles and Applications of Decision Analysis*, Strategic Decision Group: Menlo Park CA.

[41] Howard, R. (1998) Downstrean decisions and dynamic modeling, *Transcript of Panel Discussion,* 1998 Seattle INFORMS Meeting, Session SE-01, October 25th.

[42] Hwang, C-L, and Yoon, K. (1981) *Multiple Attribute Decision Making: Methods and Applications*, Springer-Verlag: Berlin.

[43] Jessop, A. (1999) Entropy in multiattribute problems, *J. Multi-Criteria Decision Making*, **8**, pp. 61-70.

[44] Kaplan, R., Norton, D. (1996) *The Balanced Scorecard*, Harvard Press: Boston.

[45] Keeney, R. and Raiffa, H. (1976) *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, Wiley: New York.

[46] Keeney, R. (1992) *Value-Focused Thinking*, Harvard Press: Cambridge.

[47] Kelman, A. and Yager, R. (1998) Using priorities in aggregation connectives, in *Aggregation and Fusion of Imperfect Information*, B. Bouchon-Meunier (Ed.), Physica-Verlag: Heidelberg.

[48] Klein, G. (1998) *Sources of Power*, MIT Press: Cambridge.

[49] Kosko, B. (1986) Fuzzy cognitive maps, *Int. J. Man, Machine Systems*, **24**, pp. 65-75.

[50] Kumar, N. and Ganesh, L. (1996) A simulation-based evaluation of the approximate and exact eigenvector methods employed in AHP, *European J. Operational Research*, **95**, pp. 656-662.

[51] Larichev, O. and Moshkovich, H. (1991) ZAPROS: A method and system for ordering multiattribute alternatives on the basis of a decision-maker's preferences, *Preprint of All Union Institute for Systems Studies*, Moscow.

[52] Larichev, O., Moshkovich, H., Mechitov, A. and Olson, D. (1993) Experiments comparing qualitative approaches to rank ordering of multiattribute alternatives, *J. Multi-Criteria Decision Analysis,* **2**, pp. 5-26.

[53] Laskey, K. and Lehner, P. (1994) Metareasoning and the problem of small worlds, *IEEE Systems, Man, and Cybernetics*, **24**(11), pp. 1643-1652.

[54] Leon, O. (1999) Value-Focused Thinking versus Alternative-Focused Thinking: Effects on Generation of Objectives, *Org. Behavior and Human Decision Processes*, **80**(3), pp. 213-227.

[55] Leskinen, P (2000) Measurement scales and scale independence in the Analytic Hierarchy Process, *J. Multi-Criteria Decision Analysis*, **9**, pp. 163-174.

[56] Lootsma, F. (1993) Scale sensitivity in the multiplicative AHP and SMART, *J. Multi-Criteria Decision Analysis,* **2,** pp 87-110.

[57] Lootsma, F. and Schuijt, H. (1997) The multiplicative AHP, SMART, and ELECTRE in a common context, *J. Multi-Criteria Decision Analysis*, **6**, pp. 185-196.

[58] Luce, R. and von Winterfeldt, D. (1994) What common ground exists for descriptive, prescriptive and normative utility theories?, M*anagement Science*, **40**(2), pp. 263-279.

[59] Marichal, J-L. (2000) An axiomatic approach of the discrete Choquet integral as a tool to aggregate interacting criteria, *IEEE Trans. Fuzzy Systems*, **8**(6), pp. 800-807.

[60] Miettinen, K. and Makela, M. (1999) Comparative evaluation of some interactive reference point-based methods for multi-objective optimisation, *J. Operational Research Society*, **50**, pp.949-959.

[61] Mikhailov, L. and Singh, M. (1999) Comparison analysis of methods for deriving priorities in the AHP, *Proc. 1999 IEEE Systems, Man, Cybernetics Vol. 1*, pp. 1037-1042.

[62] Olson, D. (1996) *Decision Aids for Selection Problems,* Springer-Verlag: New York.

[63] Olson, D., Mechitov, A. and Moshkovich (1999) Comparison of MCDA paradigms, in *Advances in Decision Analysis,* N. Meskens and M. Roubens (Eds.), Kluwer Academic: Dordrecht.

[64] Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems,* Morgan Kaufmann: San Francisco.

[65] Poh, K-L., Fehling, M. and Horvitz, E. (1994) Dynamic construction and refinement of utility-based categorization models, *IEEE Trans. Systems, Man, Cybernetics,* **24**(11), pp. 1653-1663.

[66] Pomerol, J-C and Barba-Romero, S. (2000) *Multicriterion Decision in Management: Principles and Practice,* Kluwer Academic: Dordrecht.

[67] Poyhonen, M., Hamalainen, R. and Salo, A. (1997) An experiment on the numerical modelling of verbal ratio statements, *J. Multi-Criteria Decision Analysis*, **6,** pp. 1-10.

[68] Poyhonen, M. and Hamalainen, R. (2001) On the convergence of multiattribute weighting methods, *European J. Operations Reserarch*, **129**, pp. 569-585.

[69] Rathmell, A. (1998) Assessing the IW threat from sub-state groups, *in Cyberwar 2.0: Myths, Mysteries and Reality*, A Campen and D. Dearth (Eds.), AFCEH International Press: Virginia.

[70] Roberts, F. (1979) *Measurement Theory with Applications to Decision Making, Utility and the Social Sciences*, Addison Wesley: New York.

[71] Roy, B (1973) La Methode ELECTRE II, Une Application au Media-Planning, *Operational Research '72,* M.Ross (Ed.), North-Holland: Amsterdam.

[72] Roy, B. (1993) Decision science or decision-aid science?, *European J. Operations Research,* **66**, pp. 184-203.

[73] Roy, B. and Vanderpooten, D. (1996) The European school of MCDA: Emergence, basic features and current works, *J. Multi-Criteria Decision Analysis*, **5**, pp. 22-38.

[74] Saaty, T. (1980) *The Analytic Hierarchy Process,* McGraw-Hill: New York.

[75] Saaty, T. (1990) Eigenvector and logarithmic least squares, *European J. Operations Research*, **48**, pp. 156-160.

[76] Saaty, T. (1996) *Decision Making With Dependence and Feedback - The Analytic Network Process,* RWS Publications: Pittsburgh.

[77] Saaty, T. and Vargas, L. (1984) Inconsistency and rank preservation, *J. Math. Psychology*, **28,** pp.205-214.

[78] Salo, A. and Hamalainen, R. (1997) On the measurement of preferences in the Analytic Hierarchy Process, *J. Multi-Criteria Decision Analysis*, **6**, pp.309-319.

[79] Schenkerman, S. (1994) Avoiding rank reversal in AHP decision support models, *European J. Operations Research,* **74**, pp. 407-419.

[80] Schenkerman, S. (1997) Inducement of nonexistent order by the Analytic Hierarchy Process, *Decision Sciences*, **28**(2), pp. 475-482.

[81] Schmidt, B (2000) Evaluation of ordinal ranking algorithms, Mitre Corp., http://www.mitre.org/support/papers.

[82] Schoner, B. and Wedley, W. (1989) Ambiguous criteria weights in the AHP, *Decision Sciences*, **20**, pp.462-474.

[83] Sterman, J. (2000) *Business Dynamics*, McGraw-Hill: New York.

[84] Stevens, S. (1946) On the theory of scales of measurement, *Science*, **103** (2684), pp. 677-680.

[85] Stewart, T. (1992) A critical study of the status of multiple criteria decision making theory and practice, *OMEGA,* **20**(5/6), pp. 569-586.

[86] Sugeno, M. (1974) *Theory of fuzzy integrals and its applications*, PhD Thesis, Tokyo Institute of Technology.

[87] Triantaphyllou, E. (2000) *Multi-Criteria Decision Making Methods: A Comparative Study*, Kluwer: Dordrecht.

[88] Takeda, E., Cogger, K. and Yu, P. (1987) Estimating criterion weights using eigenvectors: A comparative study, *European J. Operations Research*, **29**, pp. 360-369.

[89] von Neumann, J. and Morganstern, O. (1947) *Theory of Games and Economic Behavior*, Princeton University Press: Princeton.

[90] Wang, M. and Yang, J. (1998) A multi-criterion experimental comparison of three multi-attribute weight measurement methods, *J. Multi-Criteria Decision Analysis*, **7**, pp.340-350.

[91] Warren, L. (1999) Strategic information synthesis by globular knowledge fusion, *Proc. Information, Decision and Control,* Adelaide 1999, pp 407-412.

[92] Warren, L. (2004) *Uncertainties in the Analytic Hierarchy Process*, DSTO technical Note TN-0597.

[93] Warren, L (2006) *On Modelling Hybrid Uncertainty in Information,* DSTO Report in Progress.

[94] Watson, S. and Buede, D. (1987) *Decision Synthesis; The principles and practice of decision analysis*, Cambridge University Press: Cambridge.

[95] Webb, G. (1996) Further experimental evidence against the utility of Occam's razor, *J. Artificial Intelligence Research,* **4**, pp. 397-417.

[96] Weber, E. and Coskunoglu, O. (1990) Descriptive and prescriptive models of decision making: Implications for the development of decision aids, *IEEE Trans. Systems, Man, and Cybernetics* , **20**(2), pp. 310-317.

[97] Wendell, R. (1974) *The Processing of Information and Structure*, Lawrence Erlbaum: New York.

[98] Yager, R. and Kelman, A. (1996) Fusion of fuzzy information with considerations for compatibility, partial aggregation, and reinforcement, *Int. J. Approximate Reasoning*, **15**, pp. 93-122.

[99] Yager, R. and Kacprzyk, J. (1997) *The Ordered Weighted Averaging Operators*, Kluwer Academic: Dordrecht.

[100] Zeleny, M. (1982) *Multiple Criteria Decision Making*, McGraw Hill: New York.

[101] Zeleny, M. (1985) *MCDM: Past Decade and Future Trends,* JAI Press.

[102] Zeleny, M. (1991) Cognitive equilibrium: A knowledge-based theory of fuzziness and fuzzy sets, *Int. J. General Systems*, **19**, pp. 359-381.

[103] Zeleny, M. (1992) An essay into a philosophy of MCDM: A way of thinking or another algorithm? C*omputers in Operations Research*, **19**(7), pp. 563-566.

[104] Zhou, S. (1996) *The Development and Evaluation of Aggregation Methods for Group Pairwise Comparison Judgments*, PhD Thesis, Portland State University.

[105] Zimmermann, H. and Zysno, P. (1980) Latent connectives in human decision making, F*uzzy Sets and Systems,* **4** (1)**,** pp 37-51.

DISTRIBUTION LIST

"As per the Research Library's *Policy on electronic distribution of official series reports* (http://web-vic.dsto.defence.gov.au/workareas/library/aboutrl/roles&policies/mission.htm) Unclassified (both Public Release and Limited), xxx-in-confidence and Restricted reports and their document data sheets will be sent by email through DRN to all recipients with Australian defence email accounts who are on the distribution list apart from the author(s) and the task sponsor(s). Other addressees and Libraries and Archives will also receive hardcopies."

Structural Uncertainties in Numerical Induction Models

Lewis Warren

AUSTRALIA

DEFENCE ORGANISATION

Task Sponsor

J521 (IO Plans) HQJOC	1 Printed

S&T Program

Chief Defence Scientist	1
Deputy Chief Defence Scientist Policy	1
AS Science Corporate Management	1
Director General Science Policy Development	1
Counsellor Defence Science, London	Doc Data Sheet
Counsellor Defence Science, Washington	Doc Data Sheet
Scientific Adviser to MRDC, Thailand	Doc Data Sheet
Scientific Adviser Joint	1
Navy Scientific Adviser	1
Scientific Adviser – Army	1
Air Force Scientific Adviser	1
Scientific Adviser to the DMO	1

Chief of Command and Control Division	Doc Data Sht & Dist List
Research Leader RLCDE	Doc Data Sht & Dist List
Head Effects-based Modelling and Analysis Group	1
Task Manager and Author : Lewis Warren	1 Printed

DSTO Library and Archives

Library Edinburgh	2 printed
Defence Archives	1 printed

Capability Development Group	
Director General Maritime Development	Doc Data Sheet
Director General Capability and Plans	Doc Data Sheet
Assistant Secretary Investment Analysis	Doc Data Sheet
Director Capability Plans and Programming	Doc Data Sheet
Chief Information Officer Group	
Director General Australian Defence Simulation Office	Doc Data Sheet
AS Information Strategy and Futures	Doc Data Sheet
Director General Information Services	Doc Data Sheet
Strategy Group	
Assistant Secretary Strategic Planning	Doc Data Sheet
Assistant Secretary Governance and Counter-Proliferation	Doc Data Sheet
Navy	
Maritime Operational Analysis Centre, Building 89/90 Garden Island Sydney NSW Deputy Director (Operations) Deputy Director (Analysis)	Doc Data Sht & Dist List
Director General Navy Capability, Performance and Plans, Navy Headquarters	Doc Data Sheet
Director General Navy Strategic Policy and Futures, Navy Headquarters	Doc Data Sheet
Air Force	
SO (Science) - Headquarters Air Combat Group, RAAF Base, Williamtown NSW 2314	Doc Data Sht & Exec Summary
Army	
ABCA National Standardisation Officer Land Warfare Development Sector, Puckapunyal	Doc Data Sheet
SO (Science) - Land Headquarters (LHQ), Victoria Barracks NSW	Doc Data Sht & Exec Summary
SO (Science) - Special Operations Command (SOCOMD), R5-SB-15, Russell Offices Canberra	Doc Data Sht & Exec Summary
SO (Science), Deployable Joint Force Headquarters (DJFHQ) (L), Enoggera QLD	Doc Data Sheet
J86 (TCS GROUP), DJFHQ	Doc Data Sheet
Joint Operations Command	
Director General Joint Operations	Doc Data Sheet
Chief of Staff Headquarters Joint Operations Command	Doc Data Sheet
Commandant ADF Warfare Centre	Doc Data Sheet
Director General Strategic Logistics	Doc Data Sheet

COS Australian Defence College	Doc Data Sheet
Intelligence and Security Group	
AS Concepts, Capability and Resources	1
DGSTA , Defence Intelligence Organisation	1
Manager, Information Centre, Defence Intelligence Organisation	1
Director Advanced Capabilities	Doc Data Sheet
Defence Materiel Organisation	
Deputy CEO	Doc Data Sheet
Head Aerospace Systems Division	Doc Data Sheet
Head Maritime Systems Division	Doc Data Sheet
Program Manager Air Warfare Destroyer	Doc Data Sheet
Guided Weapon & Explosive Ordnance Branch (GWEO)	Doc Data Sheet
CDR Joint Logistics Command	Doc Data Sheet
OTHER ORGANISATIONS	
National Library of Australia	1
NASA (Canberra)	1
UNIVERSITIES AND COLLEGES	
Australian Defence Force Academy	
Library	1
Head of Aerospace and Mechanical Engineering	1
Hargrave Library, Monash University	Doc Data Sheet

OUTSIDE AUSTRALIA

INTERNATIONAL DEFENCE INFORMATION CENTRES	
US Defense Technical Information Center	1
UK Dstl Knowledge Services	1
Canada Defence Research Directorate R&D Knowledge & Information Management (DRDKIM)	1
NZ Defence Information Centre	1
ABSTRACTING AND INFORMATION ORGANISATIONS	
Library, Chemical Abstracts Reference Service	1
Engineering Societies Library, US	1
Materials Information, Cambridge Scientific Abstracts, US	1
Documents Librarian, The Center for Research Libraries, US	1
SPARES	5 Printed

Total number of copies: 34 Printed: 10 PDF: 24

Page classification: UNCLASSIFIED

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION DOCUMENT CONTROL DATA					1. PRIVACY MARKING/CAVEAT (OF DOCUMENT)	
2. TITLE Structural Uncertainties in Numerical Induction Models				3. SECURITY CLASSIFICATION (FOR UNCLASSIFIED REPORTS THAT ARE LIMITED RELEASE USE (L) NEXT TO DOCUMENT CLASSIFICATION) Documen (U) Title (U) Abstract (U)		
4. AUTHOR(S) Lewis Warren				5. CORPORATE AUTHOR DSTO Defence Science and Technology Organisation PO Box 1500 Edinburgh South Australia 5111 Australia		
6a. DSTO NUMBER DSTO-TR-1895		6b. AR NUMBER AR-013-699		6c. TYPE OF REPORT Technical Report	7. DOCUMENT DATE July 2006	
8. FILE NUMBER 2005/1039988	9. TASK NUMBER JTW 05/147		10. TASK SPONSOR COS HQJOC		11. NO. OF PAGES 50	12. NO. OF REFERENCES 105
13. URL on the World Wide Web http://www.dsto.defence.gov.au/corporate/reports/DSTO-TR-1895.pdf					14. RELEASE AUTHORITY Chief, Command and Control Division	
15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT *Approved for public release* OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE, PO BOX 1500, EDINBURGH, SA 5111						
16. DELIBERATE ANNOUNCEMENT No Limitations						
17. CITATION IN OTHER DOCUMENTS Yes						
18. DSTO RESEARCH LIBRARY THESAURUS Decision analysis, Decision theory, Modelling, Uncertain systems						

19. ABSTRACT

This report delineates a number of ways in which the results of numerical induction models, which aggregate lower level measures into meta-measures for decision making, can be unnecessarily compromised. Examples of numerical induction models include complex models for performance evaluation, measures of effectiveness synthesis, and for strategic decision analysis. A framework is proposed for identifying different types of modelling uncertainty that may be present and several of these uncertainties are discussed in detail. Some popular decision analysis techniques are also analysed highlighting any features that may introduce unnecessary uncertainty into the results. The purpose of describing these potential pitfalls is to reduce the structural uncertainty forms that may be unwittingly added to the uncertainties that already exist in the input information leading to outputs that are more meaningful. More meaningful outputs should then naturally result in improved decisions when such models are applied to Defence problems.

Page classification: UNCLASSIFIED